

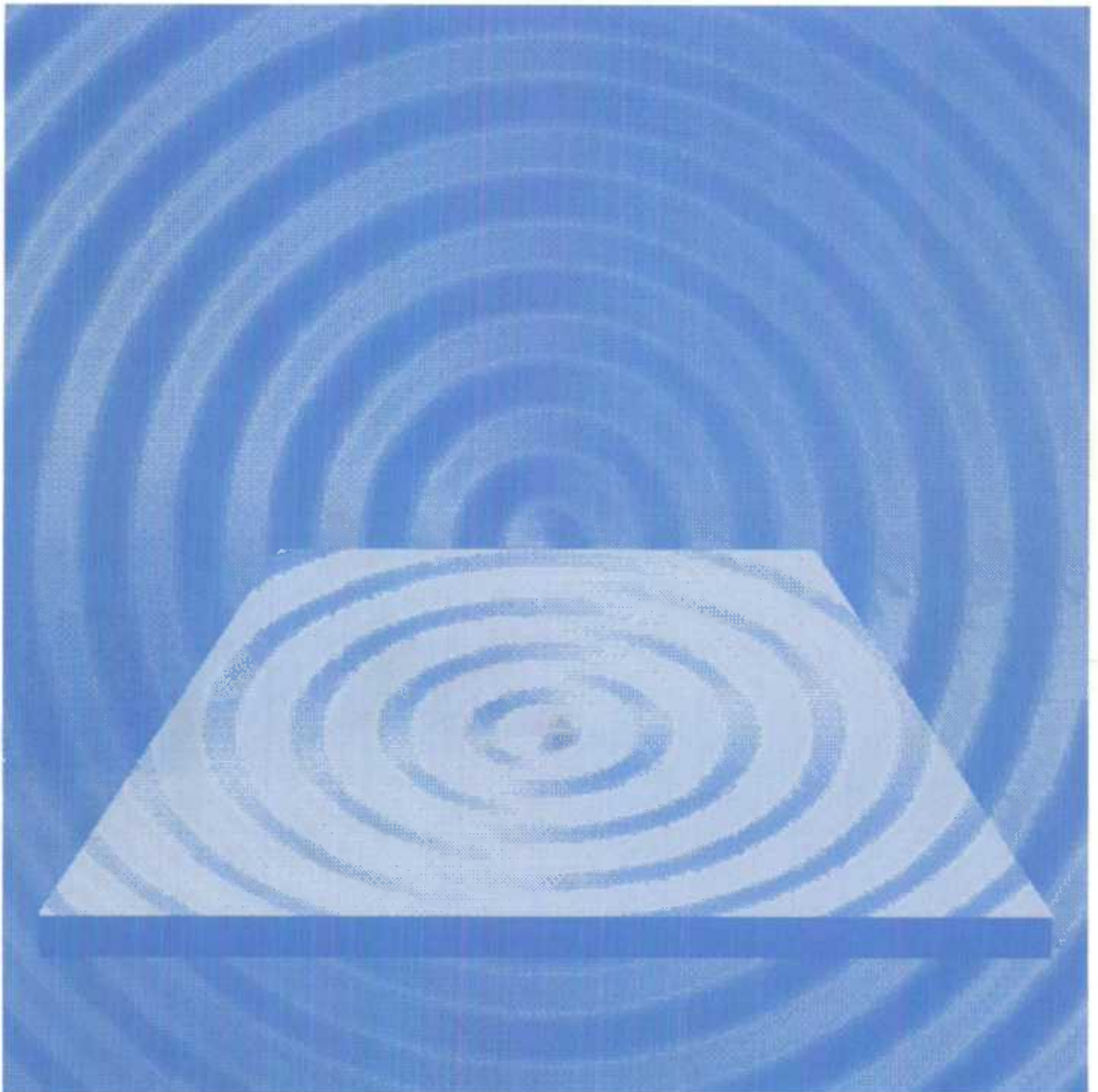
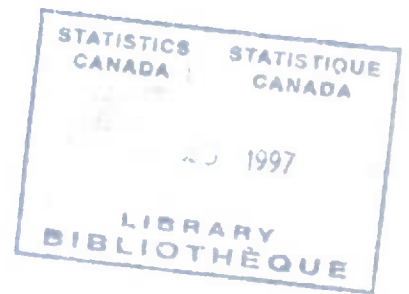


N° 11-522-XPF au catalogue

SYMPOSIUM 96

Erreurs non dues à l'échantillonnage

RECUEIL



Statistique Canada
Statistics Canada

Canada

Des données sous plusieurs formes

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur Internet, disque compact, disquette, imprimé d'ordinateur, microfiche et microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordinolingue et le système d'extraction de Statistique Canada.

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet de la présente publication ou au sujet de statistiques ou de services connexes doit être adressée à : Division des méthodes d'enquêtes-entreprises, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : (613) 951-8615) ou à l'un des centres de consultation régionaux de Statistique Canada :

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

Vous pouvez également visiter notre site sur le Web : <http://www.statcan.ca>

Un service d'appel interurbain sans frais est offert à tous les utilisateurs qui habitent à l'extérieur des zones de communication locale des centres de consultation régionaux.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Numéro pour commander seulement (Canada et États-Unis)	1 800 267-6677

Renseignements sur les commandes et les abonnements

Les prix n'incluent pas la taxe de vente

Le produit n° 11-522-XPF au catalogue est publié sur **version papier** au coût de 55 \$ au Canada. À l'extérieur du Canada, le coût est de 55 \$ US.

Faites parvenir votre commande à Statistique Canada, Division des opérations et de l'intégration, Gestion de la circulation, 120, avenue Parkdale, Ottawa (Ontario), K1A 0T6 ou commandez par téléphone au **(613) 951-7277** ou au **1 800 700-1033**, par télécopieur au **(613) 951-1584** ou au **1 800 889-9734** ou par Internet : order@statcan.ca. Pour tout changement d'adresse, veuillez fournir votre ancienne et votre nouvelle adresse. Vous pouvez aussi vous procurer les publications de Statistique Canada auprès des agents autorisés, des librairies locales et des bureaux locaux de Statistique Canada.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois et dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec le centre de consultation régional de Statistique Canada le plus près de chez vous.



Statistique Canada
Direction de la méthodologie

SYMPOSIUM 96

Erreurs non dues à l'échantillonnage

RECUEIL

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 1997

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Septembre 1997

N° 11-522-XPF au catalogue

Périodicité : hors-série

ISSN 0-660-95677-2

Ottawa

This publication is available in English upon request (Catalogue no. 11-522-XPE).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Données de catalogage avant publication (Canada)

Symposium 96, Erreurs non dues à l'échantillonnage
(1996 : Ottawa, Ont.)

Symposium 96, Erreurs non dues à l'échantillonnage :
recueil

Publié aussi en anglais sous le titre: Symposium 96, Nonsampling
errors : proceedings.

ISBN 0-660-95677-2

CS11-522-XPF

1. Statistique -- Congrès.
2. Statistique -- Méthodologie -- Congrès.
3. Services statistiques -- Qualité -- Contrôle -- Congrès.
4. Erreurs, Théorie des -- Congrès. I. Statistique Canada.
Direction de la méthodologie. II. Titre.

HA12 S9214 1997 001.4'22

C97-988023-8

Le papier utilisé dans la présente publication répond
aux exigences de l' "American National Standard for
Information Sciences" - "Permanence of Paper for
Printed Library Materials", ANSI Z39, 48 1984.



AVANT-PROPOS

Le Symposium de 1996 était le treizième de la série des symposiums internationaux sur les questions de méthodologie parrainée par Statistique Canada. Chaque année, le symposium porte sur un thème particulier. Celui de 1996 portait sur les erreurs non dues à l'échantillonnage.

Le symposium a réuni plus de 300 personnes pendant trois jours au Centre de conférences Simon-Goldberg à Ottawa. On y a entendu des présentations de statisticiens du milieu universitaire et d'agences gouvernementales et également du secteur privé. En tout, 29 communications ont été présentées par des conférenciers invités. Mis à part la traduction et le travail de mise en page, ce recueil contient les présentations telles que soumises par les auteurs.

Les organisateurs du symposium de 1996 tiennent à exprimer leurs remerciements aux nombreuses personnes qui ont contribué à la réalisation de cet ouvrage ainsi qu'à toutes celles qui les ont aidé lors de la tenue du Symposium en novembre. En particulier nous tenons à remercier Sophie Arsenault, Sophie Dionne, Guylaine Dubreuil, Hugh Gough, Guido, Josée Morel, Peggy O'Neill, Micheline Sabourin et Michelle Simard pour leur assistance lors de la préparation du matériel et des préparatifs reliés au Symposium de 1996.

Il convient, naturellement, de remercier les conférenciers pour avoir pris le temps de mettre leurs idées par écrit. La publication de ce recueil a aussi nécessité le concours de nombreuses autres personnes. Le traitement des manuscrits a été exécuté habilement par Lynn Savage. La correction des épreuves a été effectuée par de nombreux méthodologistes, dont René Boyer, Danielle Lebrasseur, Harold Mantel, Carole Morin, Claude Poirier, Michelle Simard, Jean-Louis Tambay, Charles Tardif et Alain Théberge.

Le quatorzième symposium annuel de Statistique Canada sera tenu dans la région d'Ottawa, du 5 au 7 novembre 1997. Le thème sera: Les nouvelles orientations pour les enquêtes et les recensements.

Comité organisateur du Symposium de 1996

Ann Brown

Jane Burgess

Johane Dufour

Éric Rancourt

Le lecteur peut reproduire sans autorisation des extraits de cette publication à des fins d'utilisation personnelle à condition d'indiquer la source en entier. Toutefois, la reproduction de cette publication en tout ou en partie à des fins commerciales ou de redistribution nécessite l'obtention au préalable d'une autorisation écrite de Statistique Canada.

LA SÉRIE DES SYMPOSIUMS DE STATISTIQUE CANADA

- 1984 - L'analyse des données d'enquête
- 1985 - Les statistiques sur les petites régions
- 1986 - Les données manquantes dans les enquêtes
- 1987 - Les utilisations statistiques des données administratives
- 1988 - Les répercussions de la technologie de pointe sur les enquêtes
- 1989 - L'analyse des données dans le temps
- 1990 - Mesure et amélioration de la qualité des données
- 1991 - Questions spatiales liées aux statistiques
- 1992 - Conception et analyse des enquêtes longitudinales
- 1993 - International Conference on Establishment Surveys (*en anglais seulement*)
- 1994 - Restructuration pour les organismes de statistique
- 1995 - Des données à l'information - méthodes et systèmes
- 1996 - Les erreurs non dues à l'échantillonnage

**LA SÉRIE DES SYMPOSIUMS INTERNATIONAUX DE STATISTIQUE CANADA
RENSEIGNEMENTS CONCERNANT LA COMMANDE DES RECUEILS**

Utilisez le bon de commande sur cette page pour commander des copies additionnelles du recueil du Symposium 96: «Erreurs non dues à l'échantillonnage». Vous pouvez aussi commander les recueils des derniers symposiums. Une fois complété, retournez le formulaire à:

RECUEIL DU SYMPOSIUM 96
STATISTIQUE CANADA
DIVISION DES MÉTHODES D'ENQUÊTES - ENTREPRISES
IMMEUBLE R.H. COATS, 3^e ÉTAGE
PARC TUNNEY
OTTAWA (ONTARIO)
K1A 0T6
CANADA

Veillez inclure le paiement avec votre commande (chèque ou mandat poste, en dollars canadiens ou l'équivalent, à l'ordre du «Receveur général du Canada». Veuillez indiquer sur votre chèque ou mandat poste: Recueil du Symposium 96).

RECUEIL DU SYMPOSIUM: NUMÉROS DISPONIBLES

1987 -	Les utilisations statistiques des données administratives - FRANÇAIS	_____ @ \$10
1987 -	Statistical Uses of Administrative Data - ANGLAIS	_____ @ \$10
1987 -	ENSEMBLE DE 1 ANGLAIS ET 1 FRANÇAIS	_____ @ \$12 L'ENSEMBLE
1988 -	Les répercussions de la technologie de pointe sur les enquêtes - BILINGUE	_____ @ \$10
1989 -	L'analyse des données dans le temps - BILINGUE	_____ @ \$15
1990 -	Mesure et amélioration de la qualité des données - FRANÇAIS	_____ @ \$18
1990 -	Measurement and Improvement of Data Quality - ANGLAIS	_____ @ \$18
1991 -	Questions spatiales liées aux statistiques - FRANÇAIS	_____ @ \$20
1991 -	Spatial Issues in Statistics - ANGLAIS	_____ @ \$20
1992 -	Conception et analyse des enquêtes longitudinales - FRANÇAIS	_____ @ \$22
1992 -	Design and Analysis of Longitudinal Surveys - ANGLAIS	_____ @ \$22
1993 -	International Conference on Establishment Surveys - ANGLAIS (disponible en anglais seulement, recueil publié aux É.-U.)	_____ @ \$58
1994 -	Restructuration pour les organismes de statistique - FRANÇAIS	_____ @ \$53
1994 -	Re-engineering for Statistical Agencies - ANGLAIS	_____ @ \$53
1995 -	Des données à l'information - méthodes et systèmes - FRANÇAIS	_____ @ \$53
1995 -	From Data to Information - Methods and Systems - ANGLAIS	_____ @ \$53
1996 -	Nonsampling Errors - ANGLAIS	_____ @ \$55
1996 -	Erreurs non dues à l'échantillonnage - FRANÇAIS	_____ @ \$55

S.V.P. AJOUTER LA TAXE SUR LES PRODUITS ET SERVICES (7%)
(Résidents du Canada seulement)

\$ _____

MONTANT TOTAL DE LA COMMANDE

\$ _____

S.V.P. INCLURE VOTRE ADRESSE COMPLÈTE AVEC VOTRE COMMANDE !

NOM _____

ADRESSE _____

VILLE _____ PROV/ÉTAT _____ PAYS _____

CODE POSTAL _____ TÉLÉPHONE _____ TÉLÉCOPIEUR _____

Pour plus de renseignements, communiquer avec John Kovar, téléphone: (613) 951-8615, télécopieur (613) 951-1462, internet kovar@statcan.ca.

ERREURS NON DUES À L'ÉCHANTILLONNAGE

TABLE DES MATIÈRES¹

ALLOCUTION D'OUVERTURE	3
G. Brackstone , Statisticien en chef adjoint, Statistique Canada	
DISCOURS PRINCIPAL	
Erreurs non dues à l'échantillonnage dans les enquêtes: La recherche d'une solution pratique	9
Robert M. Groves , University of Michigan	
SESSION 1: LA GESTION D'ENQUÊTE ET LES ERREURS NON DUES À L'ÉCHANTILLONNAGE	
Président: Scott Murray , Statistique Canada	
Diffusion de la qualité à Statistics Sweden: Principe et pratique	21
Claes Anderson , Håkan L. Lindström et Lars Lyberg, Statistics Sweden	
Le recensement du millénaire - un produit à qualité totale?	23
Graham Jones , Office for National Statistics, Royaume-Uni	
Normes et lignes directrices concernant l'erreur non due à l'échantillonnage: Fournir un meilleur service à moindre prix	33
Richard D. Burgess , Statistique Canada	
SESSION 2: ERREURS LIÉES À LA BASE DE SONDAGE	
Président: Larry Swain , Statistique Canada	
Mesure des erreurs du Registre des entreprises	41
Normand Laniel , Lenka Mach, Hugh Finlay et Sophie Dionne, Statistique Canada	
L'estimation de l'erreur de couverture lors du recensement de la population de 1996	55
Claude Julien , Statistique Canada	
Quel est le rôle de l'analyse démographique dans le recensement de 2000 des États-Unis?	65
J. Gregory Robinson , U.S. Bureau of the Census	

¹ Dans le cas de plusieurs auteurs, le nom du présentateur est imprimé en caractères gras.

SESSION 3: MAXIMISATION DES TAUX DE RÉPONSE

Présidente: Maryanne Webber, Statistique Canada

Le taux de réponse et l'Enquête sur la population active canadienne: Fruit du hasard ou bonne planification?	75
Mike Sheridan , Doug Drew et Benoit Allard, Statistique Canada	
Méthodes de rechange lors de l'élaboration de stratégies de communications: Augmenter les taux de réponse ou les taux de non-réponse	85
Scott D. Nowlan , Price Waterhouse	
Favoriser la réponse aux enquêtes sur l'agriculture	95
Patricia Whitridge , Statistique Canada	
Minimisation de la non-réponse dans une enquête par panel	103
Heather Laurie , Rachel Smith et Lynne Scott, University of Essex	

SESSION 4: CONCEPTION DE QUESTIONNAIRE ET CONTRÔLE DE LA QUALITÉ

Présidente: Marlene Levine, Statistique Canada

Réduction des erreurs non dues à l'échantillonnage par la recherche cognitive dans le supplément de la Current Population Survey sur la race et l'ethnicité	119
Ruth B. McKay , Bureau of Labor Statistics, É.-U.	
Mesure de la qualité durant le traitement des données d'enquête	131
Kathryn Williams , Connie Denyes et Mary March, Statistique Canada	
Contrôle des interviews téléphoniques assistées par ordinateur au U.S. Bureau of the Census	143
Mary Ellen Beach , Jane Woods et Geraldine Burt, U.S. Bureau of the Census	

SESSION 5: PROBLÈMES POSÉS PAR LES NOUVELLES MÉTHODES DE COLLECTE ET SAISIE

Président: Jean-François Gosselin, Statistique Canada

Expériences de Statistique Canada avec la saisie automatisée de données	155
Suzanne M. Vézina , Statistique Canada	
Erreur non due à l'échantillonnage: La déclaration électronique peut-elle aider?	163
Laurie Hill , Statistique Canada	
L'utilisation de W3 comme méthode de collecte de données: Conception des enquêtes de l'avenir	167
Richard L. Clayton , Bureau of Labor Statistics, É.-U.	

SESSION 6: ERREURS DE RÉPONSE

Président: M.P. Singh, Statistique Canada

L'appariement de micro-données: un outil d'évaluation et d'amélioration de la qualité des données d'enquêtes	183
Lizette Gervais-Simard , Statistique Canada	
Validité de l'auto-déclaration des problèmes de santé chroniques lors de l'Enquête nationale sur la santé de la population	195
Gary Catlin, Karen Roberts et Susan Ingram, Statistique Canada	
Enquête sur le tabagisme au Canada	197
Lecily Hunter , Statistique Canada	

SESSION 7: ERREURS DE MESURE

Président: Garnett Picot, Statistique Canada

Erreur due à l'évaluateur lors de l'évaluation du rendement de l'intervieweur	209
Paul P. Biemer , Research Triangle Institute, É.-U.	
Causes de la non-réponse à une question liée au temps, au répondant et à l'intervieweur dans l'échelle d'évaluation de l'état dépressif CES-D: un modèle multiniveau	211
Pieter van den Eeden, Johannes H. Smit et Aart-Jan Beekman, Vrije Universiteit	
Évaluation de l'erreur non due à l'échantillonnage dans les données d'enquête à l'aide de modèles avec ordonnée à l'origine aléatoire	223
Dale Atkinson , U.S. Department of Agriculture	

SESSION 8: DONNÉES ADMINISTRATIVES

Président: Geoff Hole, Statistique Canada

Quels sont les impacts sur la qualité des données lorsque des données d'enquête sur le revenu sont combinées à des données administratives?	237
Sylvie Michaud et Michel Latouche, Statistique Canada	
L'assurance de la qualité du Registre Canadien du cancer	249
Leslie A. Gaudette, Tony Labillois, Ru-Nie Gao et Heather Whittaker, Statistique Canada	
Incidence de la réglementation et des pratiques administratives sur le fonctionnement d'un système statistique axé sur un registre - Cas du registre des prestataires de l'assurance-chômage	261
Timo Koskimaki , Statistics Finland	

DISCOURS PRINCIPAL DE CLÔTURE

Président: J.N.K. Rao, Carleton University

Erreurs non dues à l'échantillonnage et estimation par sondage	277
Wayne A. Fuller , Iowa State University	

ALLOCUTION DE CLÔTURE

David A Binder, Directeur, Division des méthodes d'enquêtes-entreprises, Statistique Canada

ALLOCUTION D'OUVERTURE

ALLOCUTION D'OUVERTURE

G. J. Brackstone¹

Bonjour à tous. Au nom de Statistique Canada, je vous souhaite la plus cordiale bienvenue à notre treizième symposium sur la méthodologie. J'espère que vous ne souffrez pas de superstition, car nous sommes aussi le treize du mois. Aujourd'hui, nous accueillons des représentants de plusieurs pays, en l'occurrence des États-Unis, du Royaume-Uni, de Suède, de Finlande, des Pays-Bas et de la République des Philippines, en plus de notre habituel contingent de Canadiens.

Pour vous mettre dans le contexte, permettez-moi de vous dire quelques mots sur l'histoire de ces symposiums, qui a commencé en 1984. Pendant les quatre premières années, nous alternions entre un petit symposium intime tenu dans cette pièce et un symposium plus imposant organisé dans des locaux plus vastes du centre-ville. En 1985, par exemple, nous avons tenu, au Centre des congrès, un symposium international sur les statistiques sur les petites régions qui a donné lieu à la parution du livre intitulé *Small area statistics*, publié chez J. Wiley. En 1987, nous nous sommes également réunis au centre-ville, à la gare centrale, pour notre symposium sur les utilisations statistiques des données administratives. Ensuite, pour les cinq années suivantes, nous sommes revenus ici pour aborder des thèmes comme le recours aux technologies pour mener des enquêtes, l'analyse des séries chronologiques, les données géographiques, la qualité des données et les enquêtes longitudinales. En 1993, nous avons jumelé notre symposium annuel à la conférence internationale sur les établissements qui se tenait à Buffalo et qui a subséquentement mené à la production d'un autre livre publié chez J. Wiley. Au cours des deux années qui ont suivi, nous sommes revenus ici mais pour traiter de sujets un peu plus vastes que la méthodologie, c'est-à-dire la restructuration pour les organismes de statistique et l'an dernier : « Des données à l'information ».

Cette année, nous avons choisi pour thème un sujet au coeur même de la méthodologie des enquêtes : Les erreurs non dues à l'échantillonnage. Bien sûr il ne s'agit pas d'une nouveauté. Il y a 25 ans, on soutenait déjà qu'il fallait s'intéresser davantage aux erreurs non dues à l'échantillonnage et qu'on accordait trop de temps aux erreurs d'échantillonnage, alors que le véritable problème était ailleurs.

Nos études sur les erreurs non dues à l'échantillonnage ont-elles progressé? Il existe bien sûr de plus en plus d'ouvrages dans ce domaine. Des milliers d'études et de rapports ont été rédigés sur toutes sortes d'erreurs non dues à l'échantillonnage associées à certaines enquêtes ou à certains recensements. Mais au fait, combien d'enquêtes ont vraiment été conçues en tenant systématiquement compte de toutes les erreurs prévisibles? Dans combien d'enquêtes mesure-t-on la qualité sans se contenter de relever les erreurs d'échantillonnage et peut-être même de simples erreurs de réponse? Si les erreurs non dues à l'échantillonnage sont si importantes, ne trompons-nous pas les utilisateurs en les avertissant seulement des erreurs d'échantillonnage? Et puis, avons-nous dépassé le stade de concevoir des applications et de faire des études de cas sur des enquêtes particulières pour atteindre celui de l'élaboration d'un cadre général, ou d'une théorie de la conception et de la mise en oeuvre d'enquêtes qui tiennent compte des erreurs non dues à l'échantillonnage ou qui servent à mesurer la qualité de l'ensemble des données d'enquêtes?

En général, je crois fermement que la réponse est positive. Nous avons fait des progrès considérables, et je sais qu'au cours des prochains jours, nous allons en apprendre beaucoup sur les travaux qui ont été accomplis jusqu'ici. Permettez-moi de dire quelques mots sur le programme de ce symposium, et d'abord

¹ Gordon J. Brackstone, Statisticien en chef adjoint, Secteur de l'informatique et de la méthodologie, Statistique Canada, 26-J, Immeuble R.H. Coats, Ottawa, Ontario, K1A 0T6.

à propos de son titre. À mon avis, l'expression « erreurs non dues à l'échantillonnage » pose deux problèmes. Premièrement, elle rassemble toutes les erreurs sauf celles qui sont attribuables à l'échantillonnage et fait un salmigondis des mécanismes qui, pris isolément, mériteraient sans doute leur propre fondement théorique. L'unification de cette mosaïque de sources d'erreurs en une seule armature n'est pas chose simple. Au fil des sessions de ce symposium, lesquelles ont tendance à porter chacune sur une seule source d'erreurs, j'espère que nous ne perdrons par de vue qu'il faut ensuite intégrer ces sources d'erreur à l'ensemble des activités de conception d'enquêtes et de mesure de la qualité.

L'expression « erreurs non dues à l'échantillonnage » présente une deuxième difficulté : elle emploie un terme qu'affectionnent particulièrement les statisticiens, soit le mot « erreur » et sa connotation négative associée à la « faute ». On a l'impression que les erreurs d'échantillonnage qui, bien évidemment, ont lieu en toute connaissance de cause et à dessein, ne sont pas introduites par mégarde, tandis que les erreurs non dues à l'échantillonnage le sont : il manque à l'enquête quelque chose qui aurait dû être fait. Je pense ici que cette distinction est trompeuse. Normalement, nous ne pouvons pas nous permettre d'éviter les erreurs non dues à l'échantillonnage en obtenant un taux de réponse de 100 % ou en éliminant toutes les erreurs de codage, pas plus que nous pouvons nous permettre d'éviter les erreurs d'échantillonnage en sondant toute la population. Pour concevoir des enquêtes et obtenir la meilleure qualité possible, il faut donc faire des compromis en tenant compte de toutes les sources d'erreurs, qu'elles relèvent de l'échantillonnage ou non.

Ce qui distingue vraiment l'erreur d'échantillonnage de l'erreur non due à l'échantillonnage, c'est la connaissance et la maîtrise que nous possédons du mécanisme de génération des erreurs. En ce qui concerne l'échantillonnage, nous concevons et maîtrisons nous-mêmes le mécanisme de génération d'erreurs. En revanche, les erreurs non dues à l'échantillonnage sont dissimulées, et nous avons du mal à les comprendre. C'est pourquoi les problèmes d'erreurs non dues à l'échantillonnage sont plus difficiles à résoudre. Il est plus tentant de les ignorer et de se préoccuper des résultats plus malléables et élégants de la théorie de l'échantillonnage. Au fil des ans, nous avons néanmoins accompli des progrès considérables. Nous pouvons maintenant mieux comprendre bon nombre des mécanismes de génération d'erreurs et, par conséquent, élaborer les outils de conception qui permettront de les éliminer ou de les maîtriser.

Reste ensuite à trouver le juste milieu entre les sources d'erreurs (d'échantillonnage ou non) et les coûts et autres contraintes imposées à la préparation d'enquêtes gouvernementales. Ce qui nous amène au thème de la conception intégrale d'enquêtes, dont nous entendrons parler, je l'espère, au cours du présent symposium.

Lors des allocutions d'ouverture de ces symposiums, j'essaie habituellement d'expliquer pourquoi le moment est venu d'aborder le sujet que nous avons choisi. Cependant les erreurs non dues à l'échantillonnage semblent être un sujet de tous les temps dont on peut constamment tirer profit. Quoi qu'il en soit, permettez-moi de vous donner quelques raisons d'actualité pour expliquer un tel choix. La réputation d'un bureau de statistique dépend de la crédibilité des données qu'il diffuse. Lorsque la crédibilité est contestée, elle l'est rarement à cause de l'échantillonnage, mais presque toujours d'une question sans lien avec lui. Une deuxième raison: les bureaux de statistique sont aux prises avec des sujets de plus en plus délicats et difficiles dans les enquêtes. Les principaux obstacles auxquels se heurtent pareilles enquêtes se rapportent habituellement à des problèmes de réponse et de non-réponse plutôt qu'à l'échantillonnage. Une troisième raison : les enquêtes longitudinales sont devenues de plus en plus populaires et posent de nouveaux défis qui ne relèvent pas de l'échantillonnage. Enfin, je pense que plus on utilise et réutilise les données de façon analytique, plus elles seront comparées à d'autres sources; il devient alors essentiel de comprendre à fond le profil qualitatif des ensembles de données. Ce sont là quelques-unes seulement des raisons pour lesquelles il faut continuer à étudier ce sujet.

Comme d'autres organismes du même genre, Statistique Canada est aux prises avec un grand nombre de problèmes en ce domaine; il est donc bon de voir que tant de gens soient venus ici pour nous faire profiter de leur expérience et de leurs connaissances et nous aider à résoudre ces problèmes.

Je ne veux pas dépasser le temps qui m'est accordé et commettre ainsi la première erreur non due à l'échantillonnage à avoir lieu pendant ce symposium; alors permettez-moi de souhaiter à nouveau la bienvenue à tous. Je vous remercie de vous être présentés à ce symposium et j'espère que ces trois jours seront intéressants et utiles. Je dis tous les ans que nous, à Statistique Canada, nous profitons de façon incalculable des exposés et des échanges qui ont lieu dans le cadre de ces symposiums. J'espère que vous jugerez tous que vous et votre organisme aurez aussi profité au maximum du temps passé ici.

DISCOURS PRINCIPAL

ERREURS NON DUES À L'ÉCHANTILLONNAGE : LA RECHERCHE D'UNE SOLUTION PRATIQUE

Robert M. Groves¹

RÉSUMÉ

Dans le présent article, nous procédons à un examen historique de la structure conceptuelle des erreurs non dues à l'échantillonnage. Cette étude nous amène à formuler les observations suivantes : a) dans toutes les sciences, le progrès se traduit par une meilleure compréhension des éléments constitutifs d'un phénomène (c'est ainsi que nous disposons aujourd'hui de théories concernant les étapes de la compréhension des répondants et leurs rapports aux erreurs de mesures); b) la mise au point de mesures pratiques d'évaluation des erreurs non dues à l'échantillonnage aux fins de la conduite des enquêtes se heurte aux points de vue divergents exprimés dans les études expérimentales et dans celles fondées sur l'observation (il nous manque un groupe d'estimateurs qui permettrait d'incorporer les diverses erreurs non dues à l'échantillonnage); c) pour progresser dans notre étude des erreurs non dues à l'échantillonnage, nous aurons besoin d'outils d'estimation fondés sur des modèles, outils qui sont actuellement étrangers à la plupart des analystes d'enquête (la plupart des paramètres estimatifs de l'erreur non due à l'échantillonnage doivent s'appuyer sur un modèle quelconque du phénomène comportemental responsable de l'erreur). En guise de conclusion, nous formulons des hypothèses sur la nature des composantes de l'erreur non due à l'échantillonnage qui pourraient se prêter à la réduction ou à des mesures par suite des découvertes actuelles et à venir.

MOTS CLÉS : Erreurs non dues à l'échantillonnage; modèles d'erreurs de mesures; ajustement tenant compte de la non-réponse.

1. SURVOL DES REMARQUES

Dans toutes les conférences de recherche auxquelles j'ai assisté, j'ai eu l'occasion d'entendre au moins trois types différents de discours principaux. Le premier type est celui où on fait un survol des sujets qui seront abordés lors de la conférence. Pour ce qui est du présent symposium, une telle tâche serait considérable compte tenu de la grande diversité des travaux présentés. Le deuxième type d'exposé propose aux participants un sujet de recherche appartenant à un domaine complètement différent, mais qui présente beaucoup des problèmes qu'ils connaissent déjà. On cherche implicitement, grâce à cette stratégie, à promouvoir de nouveaux travaux, adoptant de nouvelles approches. Les exposés du troisième type abordent une découverte particulière dans le domaine d'intérêt des participants pour inciter ces derniers à se pencher sur les répercussions possibles dans le déroulement futur de leurs propres recherches.

Je m'attacherai à décrire un ensemble de résultats expérimentaux appartenant au domaine des sciences sociales, obtenus au cours des dernières années, et qui pourraient avoir une incidence sur l'approche que pourront adopter les statisticiens d'enquêtes lorsqu'ils aborderont la mise au point des modèles décrivant les erreurs non dues à l'échantillonnage. Vues sous cet angle, mes remarques pourraient faire partie du deuxième type de discours principal décrit plus haut. En choisissant cette voie, j'espère également montrer que le domaine des erreurs non dues à l'échantillonnage risque peu de se prêter à une approche théorique unique. Il nous paraît maintenant clair que les principes qui sous-tendent la couverture, la non-réponse, les effets de l'intervieweur, les effets des questions, les effets du mode et l'incidence des répondants sur l'erreur de mesure appartiendront à toute une gamme de théories des sciences sociales. De ce fait, l'expression «erreurs non dues à l'échantillonnage» convient de moins en moins pour traduire exactement les sources ou les incidences de ces erreurs. En fait, les succès remportés au cours des récentes années dans l'élucidation de ces erreurs sont dus à une dissection du processus

¹ Robert M. Groves, *University of Michigan et Joint Program in Survey Methodology*, 1218 Lefrak Hall, College Park, MD 20742.

d'enquête qui a permis d'en isoler les composantes responsables de causes différentes.

Il convient d'abord de rappeler que la recherche sur les erreurs d'échantillonnage et sur les erreurs non dues à l'échantillonnage ont des racines intellectuelles différentes. Contrairement à la recherche sur les erreurs non dues à l'échantillonnage, les statistiques d'échantillonnage s'intéressent beaucoup plus à la variance qu'au biais. Elles s'appuient sur des déductions fondées sur une théorie de la probabilité bien élaborée, par opposition à la génération d'hypothèses et à l'expérimentation. Historiquement, elles ont tendance à dominer plus nettement dans les organismes gouvernementaux de statistiques que dans les établissements de recherche universitaires.

Les théories de l'erreur non due à l'échantillonnage doivent expliquer les comportements humains responsables des erreurs. Ces théories sont surtout issues du domaine des sciences sociales. De ce fait, elles ne sont pas explicitement liées aux estimations d'enquête. Cette situation a engendré un décalage entre les découvertes des études expérimentales sur les erreurs non dues à l'échantillonnage d'une part, et leur utilisation pour la conception des enquêtes et l'estimation d'autre part. C'est l'établissement de nouveaux liens entre les deux qui nous permettra de progresser.

2. ANTÉCÉDENTS DE LA RECHERCHE

Figure 1. Articles et ouvrages portant sur des sous-domaines des enquêtes statistiques

Couverture

1936 - article de Stephan sur les problèmes des bases de sondage

Échantillonnage

1934 - article de Neyman sur la répartition en échantillonnage stratifiée

1950 - Deming, *Some Theory of Sampling*

Non-réponse

1944 - article de Hilgard et Payne sur l'absence de contacts

1946 - article de Hansen et Hurwitz sur la non-réponse

1983 - Madow et coll., *Incomplete Data in Sample Surveys*

1987 - Goyder, *The Silent Minority*

Effets de l'intervieweur

1929 - article de Rice sur les effets de l'attitude de l'intervieweur sur les réponses recueillies

1946 - article de Mahalanobis sur les échantillons superposés

1954 - Hyman, *Interviewing in Social Research*

Effets de la question

1941 - expérience de Rugg sur les effets de l'énoncé de la question

1951 - article de Hansen et de ses collaborateurs sur l'erreur de réponse

1951 - Payne, *The Art of Asking Questions*

Effets du mode

1952 - expérience de Larson sur les effets du mode

1979 - Groves et Kahn, *Surveys by Telephone*

La figure 1 fournit un bref aperçu de la chronologie des travaux de recherche portant sur les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage. On y énumère, pour chacun des types d'erreurs importants, les premiers articles publiés sur la question et la première analyse détaillée, sous forme de livre, abordant la question du double point de vue des statistiques et des sciences sociales. On constate que beaucoup des travaux précoces ont été réalisés au cours des années 1930 et 1940. La couverture et l'échantillonnage sont des domaines qui semblent appartenir exclusivement à la statistique. Par contre, la non-réponse, les effets de l'intervieweur, les effets de la question et les effets du mode ont attiré l'attention des chercheurs des deux camps.

En général, les travaux de nature statistique ont fait la promotion de modèles décrivant la nature des propriétés des estimateurs traditionnels liées à la variance, en présence de la source d'erreur particulière

Au cours des 20 dernières années, les méthodes de recherche portant sur les erreurs non dues à l'échantillonnage ont beaucoup évolué. Elles sont passées de l'étape de l'observation à celle de la planification d'expériences. Les chercheurs ont délaissé peu à peu l'étude des différences pour s'intéresser davantage à l'étude des théories sous-jacentes. Il est ainsi devenu de plus en plus évident que différents types d'erreurs non dues à l'échantillonnage exigent un recours à des théories différentes.

Il existe encore toutefois de grandes différences entre les théories portant sur les erreurs d'échantillonnage et celles portant sur les erreurs non dues à l'échantillonnage. Les premières cherchent à déterminer les circonstances en vertu desquelles on peut limiter les erreurs à un niveau particulier. L'auteur du plan d'enquête est assuré de limiter l'erreur d'échantillonnage à un niveau précis si certaines conditions présumées sont réalisées. En bref, les théories sont liées directement aux mesures de l'erreur. Par contre, dans le domaine des erreurs non dues à l'échantillonnage, la théorie a pour rôle de déterminer les causes des comportements. Beaucoup de ces théories ne se prêtent pas à un traitement mathématique. Ainsi, elles permettent de déterminer des ensembles de circonstances en vertu desquelles les erreurs peuvent être réduites, sans toutefois pouvoir quantifier cette réduction. Certaines théories proposent des méthodes de mesures fondées sur des modèles, mais la plupart ne fournissent aucun moyen de mesurer l'erreur.

Pourquoi est-il si important de pouvoir mesurer? La recherche d'un compromis entre le coût et l'erreur est un élément inhérent à toute amélioration des plans de sondage. Comme le coût est un facteur éminemment mesurable et compréhensible (même pour les députés et les *congressmen*), on a tendance à accorder moins d'attention aux erreurs commises lors des enquêtes, à moins de pouvoir en fournir une mesure quantitative. Les décisions concernant l'élaboration de la base de sondage sont quantitatives par nature, et les erreurs qui sont quantifiées attirent plus l'attention que celles qui ne le sont pas. Les erreurs non liées à l'échantillonnage ont besoin de mesures statistiques pour être prises en compte à l'étape de la conception du plan de sondage.

3. CLASSES DE THÉORIES SUR L'ERREUR NON DUE À L'ÉCHANTILLONNAGE QU'ON POURRAIT INTÉGRER DANS L'ESTIMATION

Pour parvenir à mieux prendre en compte les erreurs non dues à l'échantillonnage dans la conception des enquêtes et l'analyse de leurs résultats, il faudra selon moi mettre au point des modèles statistiques capables de mieux décrire les incidences de ces erreurs sur les estimations. De tels modèles ne pourront toutefois être utiles que s'ils intègrent les théories décrivant les comportements humains à l'origine des erreurs. Autrement dit, les théories des sciences sociales doivent guider l'élaboration du modèle, mais les statistiques doivent intégrer les modèles dans le processus d'estimation et de déduction des enquêtes.

Au point où en est actuellement la recherche sur les erreurs non dues à l'échantillonnage et les erreurs d'échantillonnage, il nous paraît possible d'envisager la modification des modèles classiques de l'erreur non due à l'échantillonnage afin d'y incorporer certains des résultats des recherches réalisées dans le secteur des sciences sociales et portant sur la non-réponse et les erreurs de mesure. On peut songer en particulier a) aux théories sur la direction du biais; b) aux théories sur les causes du biais, utiles pour l'ajustement du biais existant; c) aux théories sur le biais dans les composantes de l'erreur, habituellement assimilé à des propriétés de la variance; d) aux théories portant sur le biais et la variance de réponse.

4. THÉORIES SUR LA DIRECTION DU BIAIS - FRÉQUENCE DES COMPORTEMENTS

Il était d'usage courant, dès le milieu des années 1950, de relever les erreurs de sondage concernant le nombre de visites déclarées chez le médecin, de visites au magasin, de livres lus et d'émissions de télévision regardées. On avait parfois affaire à un biais, résultant d'une surdéclaration systématique, mais parfois aussi à une sous-déclaration pour certaines questions. À la fin des années 1980, nos connaissances sur ce phénomène avaient beaucoup progressé. Nous connaissons maintenant beaucoup mieux les incidences des questions fermées sur la compréhension, sur la tendance à recourir à l'estimation dans le cas d'événements fréquents ou réguliers (qui risque de donner lieu à une surdéclaration), et sur la tendance à recourir au dénombrement pour les événements rares ou irréguliers (qui présente un risque de sous-déclaration).

À titre d'exemple, nous reproduisons au tableau 1 les résultats obtenus par Schwarz et coll. (1985) dans une expérience avec échantillon fractionné réalisée dans le cadre d'une enquête. Les membres des deux sous-échantillons formés par sélection au hasard ont été interrogés sur leurs habitudes d'écoute de la télévision à l'aide d'une question fermée. On a offert aux membres du premier groupe un choix de six catégories de réponses allant de moins d'une demi-heure à plus de deux heures et demie, la catégorie médiane étant de une heure à une heure et demie. Les membres du second groupe avaient à choisir entre six catégories allant de moins de deux heures et demie à plus de quatre heures et demie, la catégorie médiane étant de trois heures à trois heures et demie. Comme le montre le tableau 1, plus de 84 % des personnes invitées à choisir dans la fourchette de périodes d'écoute réduites et 62 % de celles questionnées avec la fourchette de périodes d'écoute plus longues ont déclaré qu'elles regardaient la télévision moins de deux heures et demie par jour. Il semble donc qu'un choix de catégories de périodes d'écoute plus longues pousse les personnes questionnées à déclarer de plus longues heures d'écoute.

Tableau 1. Nombres déclarés d'heures d'écoute de la télévision répartis par catégories de périodes courtes ou longues (d'après Schwarz et coll., 1985)

Choix de périodes courtes		Choix de périodes longues	
	Pourcentage		Pourcentage
< 0,5 h	7 %		
0,5 - 1,0 h	18		
1,0 - 1,5 h	26	< 2,5 h	62 %
1,5 - 2,0 h	15		
2,0 - 2,5 h	18	2,5 - 3,0 h	23
		3,0 - 3,5 h	8
> 2,5 h	16	3,5 - 4,0 h	5
		4,0 - 4,5 h	2
		> 4,5 h	0
Total	100 %		100 %

Cette expérience et d'autres semblables tendent à confirmer la théorie selon laquelle le processus de formation de la réponse dans les enquêtes portant sur la fréquence d'un comportement pourrait ne pas comporter le dénombrement délibéré des événements. Au lieu de cela, dans beaucoup de cas, les répondants utilisent le choix de catégories de réponse pour porter leur jugement. Une heuristique commune utilisée pour évaluer les catégories de réponses consiste à présumer que la catégorie médiane décrit la tendance centrale de la population. Partant de cette hypothèse, les répondants sont en mesure d'indiquer qu'ils se placent dans la moyenne, sous la moyenne ou au-dessus de la moyenne et portent donc leur jugement en conséquence. Ainsi, la valeur attribuée à la catégorie médiane influe sur leur réponse.

Cette situation a de toute évidence des répercussions pratiques pour les concepteurs de questionnaires. Par exemple, compte tenu de ces résultats, on pourra décider d'éviter les questions fermées à cause de la distribution implicite de la population que le choix des catégories suppose. Toutefois, les recherches ont également démontré que les questions ouvertes portant sur la fréquence d'une activité particulière inciteront certains répondants à

fournir une réponse estimative, tandis que d'autres feront un décompte exact (sans révéler leurs stratégies de réponses). Lorsqu'on demande à des répondants d'indiquer combien de fois ils s'adonnent à une activité d'un caractère rare ou sporadique, ces derniers ont tendance à dénombrer les cas admissibles. Le risque d'erreur le plus vraisemblable, dans un tel cas, sera la sous-estimation consécutive à un oubli. Lorsque les répondants sont priés d'indiquer avec quelle fréquence ils s'adonnent à une activité d'un caractère commun et répétitif, ils ont tendance à fournir un chiffre estimatif obtenu de diverses façons. L'erreur typique, dans un tel cas, sera la surestimation due à une application partielle des règles utilisées pour obtenir cette estimation.

Quelle devrait être, dans ce cas, l'attitude du statisticien? Désignons par

Y_{e_i} = le nombre estimatif d'occasions pour la $i^{\text{ème}}$ personne;

Y_{c_i} = le nombre calculé d'occasions pour la $i^{\text{ème}}$ personne;

Y_i = le nombre réel d'occasions pour la $i^{\text{ème}}$ personne.

Selon la théorie exposée ci-dessus, $E_i(Y_{e_i}) > Y_i$ et $E_i(Y_{c_i}) < Y_i$. On obtient alors pour $0 < \theta < 1$:

$$|\text{Biais}(\theta Y_{e_i} + (1 - \theta) Y_{c_i})| < |\text{Biais}(Y_{e_i})|$$

$$|\text{Biais}(\theta Y_{e_i} + (1 - \theta) Y_{c_i})| < |\text{Biais}(Y_{c_i})|$$

Ceci pourrait signifier que pour améliorer l'estimation d'une moyenne ou d'un total Y , l'intervieweur devra peut-être poser deux questions, la première cherchant à déterminer la fréquence du comportement étudié au cours d'une semaine ou d'un mois donné, et la seconde cherchant à établir le nombre total de ces comportements au cours de la période de référence. Avec ces deux questions, on pourra établir l'estimateur mixte illustré ci-dessus avec des propriétés de biais plus intéressantes que celles obtenues avec un estimateur à une seule question.

5. THÉORIES SUR LES CAUSES DU BIAIS - RÉDUCTION DU BIAIS À L'AIDE D'UN MODÈLE D'AJUSTEMENT FONDÉ SUR LA THÉORIE

La méthode traditionnelle d'ajustement consécutif à l'enquête pour la prise en compte de la non-réponse à une question comprend les ajustements des classes de pondération et les ajustements du modèle de propension. Avec l'une ou l'autre de ces méthodes, les conditions abstraites de la réduction du biais peuvent être déduites à partir de la théorie statistique. Pour aider à déterminer si les conditions voulues seront respectées, il faut faire appel aux théories des sciences sociales.

Examinons l'ajustement classique par classes de pondération avec non-réponse dont on n'a pas à tenir compte. On présume qu'il existe J cellules d'ajustement au sein desquelles la participation est indépendante de (Y, I) , où Y désigne la variable de sondage et I la vraisemblance de l'inclusion. Dans un tel cas, l'estimateur Horvitz-Thompson de la moyenne peut prendre la forme suivante :

$$\frac{\sum \sum y_i \pi_i^{-1} r_j^{-1}}{\sum \sum \pi_i^{-1} r_j^{-1}}$$

où r_j désigne le taux de réponse pour la population de la cellule d'ajustement j .

En pratique, on peut estimer r_j en estimant d'abord la propensité à répondre à partir des données de l'échantillon. Désignons par X un caractère observé tant pour les répondants que pour les non-répondants et

$$R \parallel (Y, I) \mid X$$

alors, $r(x_j) = Pr(R_i = 1 \mid x_j)$.

La valeur de $r(x_j)$ peut être estimée indirectement à l'aide de modèles logit. On peut alors créer J cellules

d'ajustement en regroupant les valeurs estimatives $r(x_i)$ en un nombre petit de catégories. Lorsqu'on établit les modèles de propensité, on cherche à découvrir les valeurs de X souhaitables et les caractéristiques de modèles utilisables pour toute une gamme d'enquêtes.

La recherche portant sur le comportement de non-réponse a montré que le phénomène comporte de nombreuses facettes. Par exemple, le non-contact et le refus sont deux sources différentes de non-réponse qui s'appuient sur des bases comportementales très différentes. En respectant la logique exposée ci-haut, désignons par R_c la probabilité de contact, et par R_p la probabilité de participation à supposer qu'il y ait contact. Nous cherchons à déterminer si

$$R_c \text{ II } (Y, I) | X$$

mais

$$R_p \text{ II } (Y, I) | Z.$$

Quelles sont les théories qui peuvent nous éclairer sur les caractéristiques des valeurs X et Z ? On pourrait notamment songer à la théorie du scénario cognitif (*cognitive script theory*, Abelson, 1981) pour décrire le processus en vertu duquel les intentions de l'intervieweur qui cherche à obtenir une entrevue seraient interprétées par le membre du ménage interviewé. On pourrait aussi songer aux théories psycholinguistiques de la compréhension, qui nous éclaireraient sur le sens donné aux mots utilisés par l'intervieweur dans le contexte d'une conversation complète portant sur l'enquête et sur la demande d'interview. On pourrait finalement songer aux théories sociologiques des effets de la classe et de la race (Goyder, 1987).

De telles théories motivent le recours à un modèle d'ajustement en deux étapes dont la première étape reflète le processus de prise de contact avec les unités d'échantillonnage et la seconde, le processus de mise en confiance des personnes contactées. Les théories proposent de nouvelles observations à recueillir auprès des répondants et des non-répondants; elles exigent le recours à des indicateurs substitutifs des causes des deux phénomènes de non-réponse. Par exemple, le modèle de non-contact comprendrait parmi les variables du côté droit un ensemble d'influences domestiques et d'indicateurs des modes d'appel de l'intervieweur. Le modèle de coopération inclurait les influences sur la participation ainsi que les corrélats de variables d'enquête clés.

6. THÉORIES SUR LE BIAIS ET LA VARIANCE DE LA RÉPONSE - EXEMPLES D'EFFETS VARIABLES ET FIXES DES INTERVIEWEURS

La documentation statistique traitant des effets de l'intervieweur sur les données d'enquête porte essentiellement sur les modèles de composantes de la variance. On modélise les différences créées par les intervieweurs dans les données d'enquête par les diverses stratégies utilisées pour poser les questions auxquelles on attribue des erreurs de réponse différentes. À l'opposé, la documentation en sciences sociales s'intéresse surtout à un ensemble d'effets fixes exercés par le sexe, la race et l'âge de l'intervieweur sur le comportement du répondant. Ces études ont laissé constater des changements systématiques du comportement des répondants au cours d'enquêtes portant sur des questions liées aux caractéristiques démographiques susmentionnées.

Pour bien comprendre et analyser correctement ces variables et les effets fixes des intervieweurs sur les données d'enquête, il convient de modifier les caractéristiques des modèles de l'effet de l'intervieweur. Par exemple, supposons que $y_j = \mu_j + \epsilon_j$, pour chacun des $j^{\text{ème}}$ répondants. Rendons compte ensuite de l'erreur induite par l'intervieweur :

$$y_{kj} = \mu_j + b_k + e_{jk}$$

où le $j^{\text{ème}}$ répondant est attribué au $k^{\text{ème}}$ intervieweur. Les hypothèses typiques formulées dans ce modèle, principalement pour faciliter l'estimation, sont :

1. $\{b_1, b_2, \dots, b_k\}$ est un échantillon aléatoire tiré d'une population infinie ou $b_k \sim \text{iid}(0, \sigma_b^2)$
2. $e_{jk} \sim \text{iid}(0, \sigma_e^2)$

3. μ_j , b_k , e_{kj} sont sans corrélation pour tous les k, j

alors

$$\text{Var}(y) = (1/n)(\sigma_\mu^2 + \sigma_b^2 + \sigma_e^2)[1 + (m-1)\rho_j] = (\text{Var}(y_{ij})/n) [1 + (m-1)\rho_j],$$

où ρ_j désigne la corrélation intraclasse des écarts de réponses au sein de chacune des tâches de l'intervieweur. Quelles sont les théories appropriées pour la nouvelle spécification de ce modèle? La plupart sont des théories sur l'influence interpersonnelle. Elles soulignent l'influence du contexte social sur la compréhension des questions. Elles relèvent que les caractéristiques observables des intervieweurs sont des outils qui aident le répondant à comprendre la question et à porter un jugement sur les réponses appropriées. Finalement, elles font observer que la race, l'âge et le sexe des intervieweurs influent sur les réponses lorsque les questions posées ont un rapport avec ces attributs.

La modification des modèles de variance de l'intervieweur doit reconnaître qu'une portion des composantes de la variance liée aux intervieweurs est correctement liée aux attributs démographiques de ces intervieweurs. Autrement dit, une partie de la variation observée d'un intervieweur à l'autre sera due à ces attributs fixes, tandis que d'autres pourront être dues à des attributs qui seront modélisés de façon plus adéquate sous forme de composantes aléatoires. Cette perspective nous oblige à porter notre attention à la population inférentielle appropriée pour l'enquête. Par exemple, souhaitons-nous limiter l'inférence aux conditions essentielles de l'enquête où il n'y aura aucune variation de l'attribut de l'intervieweur en question? Autrement dit, souhaitons-nous mesurer la variabilité de l'intervieweur au sein d'un ensemble donné d'intervieweurs utilisés pour le projet? (L'attribut démographique est-il une variable de stratification du modèle de sélection pour les intervieweurs?) Dans l'affirmative, les modèles appropriés pourront être des modèles de variance stratifiés reflétant les effets «fixes» des attributs de l'intervieweur.

7. THÉORIES SUR LE BIAIS ET LA VARIANCE DE RÉPONSE - DATATION DES ÉVÉNEMENTS

Il existe une abondante documentation dans le domaine de la recherche-sondage sur la déclaration des dates des événements à l'intérieur de périodes de référence précisées dans le questionnaire (p. ex., « au cours des six derniers mois, quand avez-vous visité votre médecin? ». On observe communément des erreurs dans la datation des événements signalés. En règle générale, seuls les événements survenus avant le début de la période de référence peuvent être placés par erreur à l'intérieur de cette période (téléscopage en aval). Certaines observations nous portent à croire que les événements marquants, lorsqu'on se les remémore, auront tendance à paraître plus récents (téléscopage en aval). La variance de réponse en ce qui concerne la datation semble décliner pour les événements plus récents (instabilité inférieure).

Certains modèles de ce comportement ont été élaborés par des psychologues qui ont étudié la question. Par exemple, Huttenlocher et coll. (1990) proposent un modèle fondé sur diverses hypothèses, et notamment : a) un nombre d'événements uniformes dans le temps; b) un taux d'erreurs de déclaration indépendant du nombre d'événements; c) un accroissement linéaire de l'écart-type de la réponse en fonction du temps. À partir de ce modèle, on peut élaborer un estimateur du nombre total d'événements, corrigé pour tenir compte du biais, dans lequel il faudra préciser la variance de réponse, le taux d'accroissement du taux d'instabilité en fonction du temps, la longueur de la période de référence et la fréquence d'apparition. De tels modèles, qui nécessitent un grand nombre d'hypothèses de départ, ne sont pas utiles pour les estimations d'enquête.

Serait-il possible d'élaborer des modèles plus pratiques? Le problème tient au fait que la théorie présente une plus grande complexité que les modèles susmentionnés. Le biais et la variance de réponse changent tous les deux en fonction de la date réelle de l'événement. On pourrait songer à élaborer un estimateur permettant une pondération différentielle des déclarations en fonction de la variance de réponse. Un tel estimateur utiliserait des indicateurs multiples de date (p. ex., rappel libre de la date, mesure du temps écoulé depuis l'événement, recours à un calendrier); une estimation de la variance de réponse par le biais d'une analyse de covariance des indicateurs multiples et de l'utilisation de la variance de réponse dans l'estimation. Toutefois, un tel estimateur ne permet pas

de prendre en compte les propriétés de biais des réponses.

Lorsque les théories précisent à la fois les propriétés du biais et de la variance de réponse, le concepteur de l'enquête doit inventer des indicateurs multiples dont les propriétés de la variance et du biais varient selon un mode connu afin d'estimer les paramètres d'un modèle de mesure qui refléteront à la fois le biais et la variance.

8. RÉSUMÉ - APPLICATION DES THÉORIES DE L'ERREUR NON DUE À L'ÉCHANTILLONNAGE AUX ESTIMATIONS D'ENQUÊTE

Les théories portant sur les erreurs observées issues du domaine des sciences sociales ont parfois des incidences sur les caractéristiques statistiques des estimateurs. En conséquence, le statisticien chargé de l'élaboration du plan de sondage, des mesures et des estimations doit se pencher sur ces théories et inventer des moyens d'en tenir compte dans son travail. Malheureusement, rares sont les spécialistes des sciences sociales intéressés aux comportements responsables d'erreurs de sondage qui cherchent à déterminer les répercussions de ces erreurs sur les estimateurs des enquêtes. Ces spécialistes s'intéressent d'abord à l'identification des principes comportementaux applicables à une large variété de contextes (y compris les enquêtes). Le comportement de non-réponse et de réponse ne constitue, de ce point de vue, qu'un exemple de ces comportements.

Il nous paraît clair aujourd'hui que le « grand pas en avant » de la recherche sur les erreurs non dues à l'échantillonnage ne pourra être franchi sans la collaboration des spécialistes de la cognition et des sciences sociales et des statisticiens. Cette collaboration mettra l'accent sur un ensemble de méthodes de mesures et d'estimateurs incorporant des modèles d'erreurs auxiliaires. La spécification des erreurs de ces estimateurs sera motivée par les théories élaborées par les spécialistes des sciences sociales au sujet des comportements à l'origine des erreurs qui influent sur la qualité des données d'enquêtes.

9. BIBLIOGRAPHIE

- Abelson, R.P. "Psychological status of the script concept," *American Psychologist*, Vol. 36, 1981, p. 715-729.
- Deming, W.E., *Some Theory of Sampling*, New York: Wiley, 1950.
- Goyder, J., *The Silent Minority*, Boulder, CO: Westview, 1987.
- Groves, R.M. et Kahn, R.L., *Surveys by Telephone*, New York: Academic Press, 1979.
- Hansen, M.H. (1951) "Response Errors in Surveys," *Journal of the American Statistical Association*, 46, 147-190.
- Hansen, M.H. et Hurwitz, W.N., "The problem of non-response in sample surveys," *Journal of the American Statistical Association*, 41, 1946, p. 517-529.
- Heneman, G.H. et Paterson, D.G., "Refusal rates and interviewer quality," *International Journal of Opinion and Attitude Research*, 3, 1949, p. 392-398.
- Hilgard, E.R. et Payne, S.L. (1944) "Those Not at Home: a Riddle for Pollsters," *Public Opinion Quarterly*, 8:2, 254-261.
- Huttenlocher, J., Hedges, L.V. et Bradburn, N.M., "Reports of Elapsed Time: Bounding and Rounding Processes in Estimation," *Journal of Experimental Psychology, Learning, Memory and Cognition*, 16, 1990, p. 196-213.
- Hyman, H., Cobb, W.J., Feldman, J., Hart, C.W. et Stember, C., *Interviewing in Social Research*, Chicago: University of Chicago Press, 1954.

- Larson, O.N., The Comparative validity of Telephone and Face-to-face Interviewers in the Measurement of Message Diffusion from Leaflets," *American Sociological Review*, 17, 1952, p. 471-476.
- Madow, W.G., Nisselson, H. et Olkin, I., *Incomplete Data in Sample Surveys*, Volumes 1-3, New York: Academic Press, 1983.
- Mahalanobis, P.C., "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," *Journal of the Royal Statistical Society*, 109, 1946, p. 325-370.
- Neyman, J., "On the Two Different Aspects of the Representative Methods: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, 109, 1934, p. 558-606.
- Payne, S.L. (1951) *The Art of Asking Questions*, Princeton, N.J.: Princeton U. Press.
- Rice, S.A. (1929) "Contagious Bias in the Interview: a Methodological Note," *American Journal of Sociology*, 35, p. 420-423.
- Rugg, D., "Experiments in Wording Questions," *Public Opinion Quarterly*, 5, 1941, p. 91-92.
- Schwarz, N., Hippler, H.J., Deutsch, B. et Strack, F., "Response Categories : Effects on Behavioral Reports and Comparative Judgments," *Public Opinion Quarterly*, 49, p. 388-395.
- Stephan, F.F., (1936) "Practical Problems of Sampling Procedure," *American Sociological Review*, 1, p. 569-580.

SESSION 1

**LA GESTION D'ENQUÊTE ET LES ERREURS NON DUES À
L'ÉCHANTILLONNAGE**

DIFFUSION DE LA QUALITÉ À STATISTICS SWEDEN: PRINCIPE ET PRATIQUE

Claes Andersson¹, Håkan L. Lindström et Lars Lyberg

RÉSUMÉ

Cette présentation décrit les principes généraux de définition de la qualité et de déclaration de la qualité à Statistics Sweden. On discute des développements des deux dernières décennies à la lumière des préoccupations des utilisateurs de statistiques. La présentation du contexte sert ensuite à expliquer de quelles façons les conditions, les techniques et les ressources ont changé les possibilités d'étudier et de mesurer la qualité. Pour préciser les composantes principales de la qualité, on définit le niveau de qualité des sous-composantes une par une. On mentionne des approches générales pour promouvoir la mesure et la présentation d'un produit de qualité. Finalement, on présente des exemples de produits ayant une diffusion de bonne qualité.

NOTE

Cet article a été présenté au "Council of Professional Association on Federal Statistics" à Bethesda, VA, les 12 et 13 novembre, 1996. Pour se le procurer voir référence.

RÉFÉRENCE

Andersson, C., Lindström, H.L et Lyberg, L. (1996), Quality Declarations at Statistics Sweden: Principle and Practice. Proceedings of the Council of Professional Association on Federal Statistics, à paraître.

¹Claes Andersson, Statistics Sweden, SCB, UASTM, S-70189 Örebro, Sweden.

LE RECENSEMENT DU MILLÉNAIRE - UN PRODUIT À QUALITÉ TOTALE?

Graham C. Jones¹

RÉSUMÉ

L'exécution du recensement étant une entreprise très coûteuse, il est essentiel de planifier et de gérer le processus dans son ensemble, de la conception du questionnaire à l'analyse et à la présentation des résultats, en passant par les méthodes de dénombrement sur le terrain et les processus de codage et de saisie des données, en vue de fournir un produit de qualité totale.

L'auteur décrit dans les grandes lignes les mesures prises au Royaume-Uni, dans le cadre de la planification du Recensement de 2001, pour réduire au minimum le sous-dénombrement et améliorer la qualité des réponses (et les taux de réponse), puis examine la possibilité d'appliquer les méthodes d'imputation neuromimétique et les progrès réalisés dans les domaines du traitement d'images et de l'autocodage pour arriver à diffuser les analyses pertinentes dans les délais annoncés pour la gamme correcte de produits.

Les textes législatifs qui régissent le recensement exigent que celui-ci soit approuvé par le Parlement. Il est donc essentiel de convaincre les ministres qu'un équilibre doit être établi entre la limitation du fardeau de réponse, la satisfaction des exigences et le traitement des questions délicates. Pour assurer le succès de l'entreprise, il sera indispensable de régler les conflits qui surviendront naturellement dans de telles circonstances et de veiller à la convergence du processus de décision et de l'objectif global.

Mots Clés: Recensement; qualité totale; planification 2001; sous-dénombrement.

1. INTRODUCTION

Il serait présomptueux de ma part de croire que je pourrais aborder toutes les questions liées au recensement du millénaire qui se produira au Royaume-Uni dans le peu de temps que j'ai à ma disposition. En effet, je ne peux réellement parler au nom du Royaume-Uni car, bien que je porte le titre de directeur du recensement à l'Office for National Statistics et que je dirige le service de statistique gouvernemental, parallèlement, je ne suis conservateur des actes de l'état civil que pour l'Angleterre et le Pays de Galles. Mes homologues d'Écosse et d'Irlande du Nord s'occuperont du recensement dans leur propre pays.

Je n'ai pas non plus l'intention de me pencher sur cette question particulière. Je me contenterai de dire que le recensement se déroule comme une activité bien orchestrée (d'aucuns préféreraient le terme «partenariat»), même s'il diffère à certains égards dans les quatre pays qui forment le Royaume-Uni. Rien de plus normal, puisque les questions sur lesquelles on se penchera varient.

2. BUTS ET OBJECTIFS

Recenser la population et les logements, traiter et analyser les données qui en résultent et diffuser des résultats, tout cela d'une manière efficace, en insistant sur l'exactitude et l'intégrité sans compromettre la vie privée des répondants. Tout cela paraît relativement simple! Néanmoins, pareille explication ne dit rien de la raison pour laquelle on entreprend un recensement au départ. Lorsqu'on se rappelle que le recensement se trouve au cœur même des statistiques officielles, joue un rôle capital dans la planification

¹ Graham C. Jones, Directeur du recensement, Office for National Statistics, Royaume-Uni.

d'un pays et représente la seule source nationale de données ayant une couverture de 100 %, alors seulement commence-t-on à aborder le problème. Si l'on songe que derrière ces explications assez simplistes se cache l'affectation de milliards de livres de fonds publics aux administrations régionales et locales ainsi qu'aux services de santé, et la diffusion d'une somme appréciable d'information en vue de la prestation de services d'éducation et de maints services locaux, on comprend mieux les motivations économiques fondamentales à l'origine du recensement. Je n'invente rien de nouveau, bien sûr, mais cette précision est à mon avis cruciale pour ce qui va suivre et pour expliquer l'approche que nous envisageons à l'égard des principales questions.

3. DÉFINIR LES BESOINS

Lorsqu'on étudie les questions qui feront partie du recensement, il est essentiel de savoir à quoi servira l'information et qui l'utilisera. J'en ai déjà donné une idée et ceux qui participent à l'exercice connaissent les sept groupes consultatifs qui alimentent le débat. Ces groupes viennent du gouvernement central, des administrations locales, des services de santé, de la collectivité universitaire (ce qui, je l'admets, suggère une certaine homogénéité) et des entreprises (où une telle homogénéité n'existe pas!), auxquels s'en ajoutent deux spécifiques représentant l'apport de l'Écosse et de l'Irlande du Nord. Ces groupes consultatifs ont eux-mêmes engendré plusieurs groupes de travail sur le contenu du recensement, dont les propositions et les suggestions ont été sérieusement examinées lors de l'essai de 1997, auquel je reviendrai bientôt. Les consultations ont fait ressortir des problèmes manifestes, outre les questions habituelles associées au recensement. Il existe désormais une demande pour l'information sur le revenu, sur la situation socioéconomique, sur la scolarité et sur d'autres aspects avec, en arrière-plan, les profonds changements observés au niveau social et technologique. Quoi qu'il en soit, le recensement doit par nécessité faire preuve d'équilibre, doit constituer un compromis entre la demande d'information et le fardeau que l'on peut raisonnablement imposer aux répondants, cela dans un cadre de planification général illustrant clairement que les ressources consacrées au recensement ne sont pas inépuisables, donc que des conflits surgiront inévitablement et qu'il faudra accepter certaines concessions. Ces dernières ne devront pas affecter la qualité du produit et cet aspect est vital, car il importe que l'utilisateur accorde la plus grande confiance aux données qu'on lui fournit.

4. LEÇONS TIRÉES DE 1991

À cet égard, on peut tirer quelques leçons du recensement de 1991 et de l'exercice d'évaluation et de réalignement de la politique du recensement entrepris en 1992. Sachant que le recensement est nécessaire et qu'il est juste de vouloir atteindre un équilibre entre le nombre de questions et le fardeau du répondant, en fonction des exigences établies, on doit par ailleurs s'assurer que c'est le bon produit qui est engendré, au degré de précision approprié, dans des délais convenables et pour la gamme voulue de produits. Les consultations dont j'ai déjà parlé joueront aussi un rôle déterminant à cet égard.

Le recensement de 1991 a donné lieu à quelques coups d'éclat qu'on ne devrait et qu'on ne peut pas ignorer. Il y a bien sûr eu le succès de la nouvelle question sur l'ethnicité, la forte couverture globale et un tout nouveau produit - l'échantillon d'enregistrements anonymes - que d'autres pays nous envient à maints égards. Tout cela montre combien nous avons oeuvré en étroite collaboration avec de nombreux utilisateurs, notamment le professeur Angela Dale, de Manchester. Toutefois, nous avons aussi éprouvé quelques problèmes, problèmes qu'on ne peut passer sous silence : un sous-dénombrement élevé dans les noyaux urbains, des réponses inadéquates aux questions et des retards dans la production des résultats. Le fait que certains résultats reposent sur un échantillon de 10 % de la population a aussi déplu à beaucoup de gens.

5. PLANIFICATION DU RECENSEMENT DE 2001

La seule façon d'aborder la planification d'un projet aussi ambitieux que le recensement consiste à le fractionner en éléments plus petits, plus maniables. Quelques-uns de ces éléments gardent néanmoins une taille appréciable et cette désagrégation en projets soulève elle-même d'autres difficultés.

Les projets essentiels à la gestion du recensement sont les suivants :

- géographie;
- collecte des données;
- saisie des données;
- stratégie des systèmes d'information;
- production des produits, politiques et commercialisation;
- exigences en matière de données;
- sous-dénombrement et divulgation;
- vérification, imputation et qualité des données;
- législation.

Le mécanisme de production des rapports sur ces différents aspects fonctionne de façon identique en ce sens que les recommandations sont transmises au Census Programme Board. Mais l'organisation du recensement n'est pas le but de cet exposé.

Le point important à souligner est qu'on se penche sur de nouvelles méthodes concernant le déroulement du recensement dans le cadre de chaque projet :

- le projet sur la géographie servira à implanter de nouveaux SIG et recourra à la cartographie numérique pour établir les districts de recensement;
- le projet de collecte des données porte sur les méthodes postales d'envoi et de renvoi utilisées pour distribuer et récupérer les formulaires, pour la première fois au Royaume-Uni;
- le projet de saisie des données se rapporte à la saisie automatique des données par balayage au moyen de techniques ROC, et introduira un logiciel de codage automatique qui permettra le traitement intégral des données;
- le projet sur la stratégie des systèmes d'information assurera la mise en place d'un système intégré, radicalement différent de la plate-forme classique de l'ordinateur principal;
- le projet sur la politique de production et de commercialisation des produits prévoit l'examen de nouveaux produits articulés sur les médias naissants. Une certaine souplesse est essentielle à cet égard - nous sommes conscients des changements technologiques qui se produisent et savons qu'il faut garder l'esprit ouvert. Nous souhaitons aussi trouver d'autres partenaires en vue d'offrir un éventail complet de produits à valeur ajoutée;
- le projet sur les exigences en matière de données concerne les nouvelles questions et un nouveau questionnaire intégrant les exigences relatives à la saisie automatique des données par balayage;
- enfin, le projet sur la législation n'a pas besoin d'explications, mais sert de pivot dans nos relations avec l'Union européenne, en prévision de l'approche commune au recensement qu'on envisage d'adopter d'ici le nouveau millénaire.

J'ai sciemment évité de résumer les principaux projets de statistique. Ils ont principalement pour but de résoudre les problèmes associés à la qualité du produit et d'aplanir les grandes difficultés issues du recensement de 1991.

6. SOUS-DÉNOMBREMENT

Au risque de me répéter, le recensement de 1991 s'est caractérisé par une excellente couverture - de près de 98 % - mais un problème considérable est venu s'y greffer en ce sens que le sous-dénombrement n'était pas réparti de façon uniforme dans l'ensemble de la population. Plusieurs sous-groupes de la population étaient fortement sous-dénombés et le sous-dénombrement se concentrait dans quelques régions. Il en résulte de sérieux problèmes pour l'usage fondamental des données de recensement au niveau infrarégional et de graves difficultés sur le plan de la répartition des ressources.

Les groupes sous-dénombés étaient principalement les suivants :

- jeunes adultes de 20 à 29 ans, plus précisément jeunes hommes des noyaux urbains où le sous-dénombrement est estimé à environ 23 %;
- enfants de moins d'un an - le sous-dénombrement a été estimé à 3 % pour l'ensemble du pays;
- personnel des forces armées et leurs personnes à charge;
- femmes âgées d'Angleterre et du Pays de Galles - soit près de 6 % des personnes de 85 ans et plus.

Si on a pu oublier quelques adresses en 1991, on estime généralement que les logements ont bien été couverts et qu'il en a été de même pour les personnes y habitant, lorsqu'il n'y avait qu'un ménage. La principale difficulté survient où il y a multiplication des occupants, c'est-à-dire où plus d'un ménage vit à la même adresse. En 2001, les problèmes de couverture devraient être au moins aussi importants que ceux relevés en 1991. En toute probabilité, le problème se sera aggravé. En effet, les ménages d'une personne et les personnes absentes pendant la journée seront plus nombreux. En outre, plus de ménages seront difficiles à recenser à cause d'un accroissement des mesures de sécurité, comme les interphones à l'entrée du bâtiment. La population mobile, c'est-à-dire ceux qui vivent à plus d'une adresse, devrait aussi s'être multipliée.

Qu'a-t-on appris de la situation qui existait en 1991?

- La publicité commerciale était insuffisante et manquait d'efficacité, car elle ne visait pas les groupes pour lesquels existait un risque élevé de sous-dénombrement.
- La publicité n'a pas abordé le problème de la résistance à l'égard des autorités.
- Il a été difficile de recruter des enquêteurs, si bien que le personnel sur le terrain était insuffisant.
- Les méthodes applicables aux occupants multiples étaient complexes et difficiles à expliquer.

- La proportion de personnes non rejointes était plus élevée que prévu - un problème auquel on n'avait pas songé.
- Le questionnaire était trop dense et les instructions manquaient de clarté.
- On a trop attendu pour établir des contacts localement.

Enfin, comme si ce n'était pas suffisant, la planification n'a pas débuté assez tôt!

Ces leçons sont douloureuses, mais on ne peut les négliger pour autant.

Qu'envisage-t-on spécifiquement à l'égard de ces problèmes pour le recensement de 2001?

Voici les points sur lesquels on se penche :

- définition de la base démographique;
- date du recensement;
- planification géographique;
- conception et contenu du formulaire;
- recrutement et formation de personnel sur le terrain;
- établissement de contacts locaux;
- amélioration de la publicité.

La définition de la base démographique ne doit pas seulement être générale - pour éviter les omissions susceptibles d'entraîner un sous-dénombrement. Elle doit aussi être comprise et communiquée à la population ainsi qu'au personnel sur le terrain.

Les étudiants seront recensés là où ils poursuivent leurs études. Le recensement devant avoir lieu durant l'année scolaire, les étudiants des universités et des collèges et les élèves des écoles primaires seront recensés à ces endroits.

On passera en revue les difficultés que soulève la définition des résidents des établissements communautaires pour s'assurer qu'ils sont correctement dénombrés, notamment les personnes âgées, car on croit que cet aspect est à l'origine du sous-dénombrement de ce groupe. Les personnes qui comptent plus d'une adresse - habituellement celles qui travaillent dans une partie du pays et vivent ailleurs - seront aussi recensées de la façon appropriée.

Nous nous efforçons d'améliorer la planification géographique en créant des cartes spéciales, en utilisant des listes d'adresses préimprimées et en prenant les données actualisées des autorités locales - éléments qui seront intégrés à l'essai de 1997. Pareilles améliorations permettront aux enquêteurs d'être conscients de certaines difficultés probables dans leur région, à l'égard du nombre d'occupants dans les établissements non résidentiels.

L'aspect et le contenu du formulaire jouent un rôle déterminant dans le taux de réponse et la qualité des données recueillies. L'apparence, l'agencement et la formulation des questions ont fait l'objet d'études qui devraient rendre les réponses plus faciles. Dans le cadre de l'essai de 1997, on prévoit une évaluation

méticuleuse des questions susceptibles de nuire à la couverture et cet exercice fera partie de l'enquête de suivi.

En améliorant le recrutement et la formation du personnel sur le terrain grâce à des modules portant spécifiquement sur les problèmes relevés en 1991 et grâce à une étroite collaboration avec les administrations locales, le service de police et les organisations communautaires, nous nous attaquerons de plein fouet aux problèmes de couverture dans les noyaux urbains.

Enfin, la publicité doit s'adresser spécifiquement aux groupes qui posent les problèmes dont j'ai déjà parlé, c'est-à-dire les jeunes adultes, les jeunes mères et les personnes âgées. Le message sera simple et dira exactement ce qu'on attend des répondants. Nous insisterons aussi sur le recensement dans le secteur de l'éducation au moyen de messages positifs en expliquant le but.

7. LE PROGRAMME DE GESTION DE LA QUALITÉ DES DONNÉES

L'introduction du Programme de gestion de la qualité des données destiné à accroître et à mesurer la qualité des données constituera une innovation en 2001. Ce programme est une facette de l'approche de la gestion de la qualité totale (GQT) qui a connu tant de succès en Australie.

La GQT suppose la genèse d'un produit adapté aux besoins de la clientèle par un contrôle de qualité des tâches entreprises partout au sein de l'organisation. Un engagement de ce genre en matière de qualité doit naître au sommet de la pyramide et chacun dans l'organisation doit bien le saisir. Voici quelques exemples des activités qu'on entreprendra en vertu d'une telle approche :

- consultation générale de la clientèle;
- développement des systèmes par voie de prototypes;
- essai du recensement de 1997.

Le Programme de gestion de la qualité des données (PGQD) comporte plusieurs volets :

- coordination dans toutes les régions du recensement de 2001 en vue de faire ressortir les problèmes de qualité des données et d'élaborer des normes à cet égard, en consultation avec les spécialistes du traitement des données;
- contrôle et mesure de la qualité des données pendant le recensement afin de permettre l'identification et la résolution des problèmes;
- collecte de renseignements pour illustrer la qualité des données diffusées à l'utilisateur, par exemple, le degré d'imputation de chaque question;
- collecte de données en vue de l'obtention d'un chiffre unique de recensement.

Un élément essentiel consiste donc à mettre en place un mécanisme permettant de contrôler et de mesurer la qualité des données durant le recensement proprement dit - en l'occurrence le Système de contrôle de la qualité des données (SCQD). Ce système nous aidera à suivre la qualité des données à partir de normes préétablies grâce aux renseignements recueillis dans le cadre des activités du recensement (collecte, saisie des données, traitement et production) et à des sources d'information indépendantes (recensements antérieurs, dossiers administratifs, etc.). Parallèlement, le personnel connaissant les domaines spécialisés pourra interroger la base de données par le biais d'installations spéciales.

On n'a pas accordé assez d'importance à la précision géographique lors du recensement de 1991. La base géographique a posé des difficultés et a montré les incohérences qui existaient entre les districts de recensement et les codes postaux. Une bonne partie de l'information venant des sources extérieures était incorrecte et les modifications ont été reçues très tard dans la phase de planification. On a donc dû apporter les corrections et des changements à des bases de données géographiques distinctes, soit celle des entrées et celle des sorties. Or, ces bases n'étaient pas toujours synchronisées, ce qu'on n'a découvert que plus tard. On aura recours à des systèmes d'information géographiques et à la cartographie numérique, ainsi qu'à des données spécifiques sur les régions - par exemple l'existence d'un pénitencier ou d'un établissement communautaire susceptible de fausser les tendances - très tôt lors du traitement.

Les définitions peu précises du lieu de résidence, à l'origine de la confusion observée en 1991, seront révisées, comme je l'ai déjà mentionné. Une mauvaise interprétation du terme «résident» explique les erreurs relevées sur le formulaire et le fait que certaines personnes n'ont pas été recensées. Les nouvelles définitions seront testées en 1997 et on insistera particulièrement sur les règles s'appliquant aux établissements communautaires et aux visiteurs.

Le Programme de gestion de la qualité des données s'attaquera aussi à la variation du sous-dénombrement, signalée précédemment. En alimentant un système d'information sur la gestion des activités sur le terrain pour déceler les problèmes de sous-dénombrement à mesure où ils se développent, on verra comment déployer les ressources pour surmonter des difficultés particulières.

Les techniques de saisie des données par balayage ne pourront qu'améliorer la qualité de l'information originale, car il n'y aura pas d'erreur de transcription au moment de la saisie, comme cela s'est déjà produit. D'autres difficultés se présenteront néanmoins, consécutivement aux réponses cochées en trop grand nombre. Le Programme de gestion de la qualité des données illustrera l'incidence du problème - c'est-à-dire identifiera les questions qui paraissent ambiguës et veillera à ce qu'on utilise de bonnes règles et techniques de correction.

Une interprétation incorrecte et incohérente des règles de codage et les erreurs relatives aux ensembles de données de référence ont entraîné l'application de codes erronés lors des recensements antérieurs. Un codage automatique nous aidera dans une large mesure à surmonter cette difficulté, mais le PGQD établira la fréquence des champs codés, ce qui permettra d'en comparer la distribution avant et après le codage géographique. On procédera aussi à des vérifications entre différents champs, pour vérifier leur cohérence. Planifier d'avance une forte concentration de certains groupes ethniques dans des quartiers précis des noyaux urbains n'est qu'un exemple d'une telle approche.

Un des problèmes sérieux cernés en 1991 était la découverte d'erreurs à l'étape de la totalisation, étape où il est trop tard pour apporter des corrections. Les ensembles de données reposant sur l'échantillon global et l'échantillon de 10 % présentaient une énorme incompatibilité, ce qu'on n'a constaté qu'après diffusion des données. En contrôlant les données en cours de traitement et en procédant sans tarder à des totalisations croisées et à des calculs de fréquence, on identifiera les difficultés beaucoup plus tôt dans le cycle.

8. IMPUTATION

En 1991, on a recouru à une méthode d'imputation séquentielle «hot deck» s'inspirant de celle de Holt et Fellegi pour les questions qui avaient échappé à la matrice de contrôle. Cependant, la principale difficulté est qu'on ne pourra pas vérifier la cohérence après imputation en 2001. Nous poursuivons des essais sur les aspects statistiques pratiques de l'application des ordinateurs neuroniques à l'imputation et, jusqu'à présent, les résultats sont encourageants. Néanmoins, nous envisageons toujours d'autres méthodes, dont une modification du système «hot deck» de 1991, une modélisation à niveaux multiples et la nouvelle technique d'imputation canadienne (NIM). Ces méthodes devront être évaluées afin qu'on établisse leur

performance. Avec les contrôles de cohérence suivant l'imputation, on devrait ainsi aplanir les principales difficultés. Le professeur Ray Chambers nous prête son concours dans le cadre de cette évaluation.

9. LE RECENSEMENT À CHIFFRE UNIQUE

Le chiffre du recensement sert de valeur de référence sur laquelle s'alignent les estimations de population annuelles. Comme je l'ai déjà mentionné, ces données sont essentielles à la répartition des fonds publics. Il est donc capital que l'utilisateur ait la plus grande confiance dans l'information fournie. Qu'on en vienne à douter des estimations régionales ou à les remettre en question, et avec elles la façon dont elles se comparent d'une région à l'autre, et c'est l'utilité même du recensement qui est compromise.

Pour cette raison, nous avons convenu qu'un des principaux objectifs de l'Office for National Statistics serait d'entreprendre des recherches et de voir si on pourrait fournir l'information du recensement à tous les niveaux, à partir d'une base commune en prévision du recensement de 2001.

L'exercice n'exigera pas seulement de l'expertise technique, mais une gestion très méticuleuse et la résolution des difficultés que soulève la production d'un ensemble de données de cette envergure, ce qu'il ne faudrait pas sous-estimer. Le projet sera dirigé par le professeur Diamond de l'Université de Southampton, qui remettra son rapport à un comité directeur présidé par le chef de la méthodologie statistique à l'Office for National Statistics. La qualité du travail sera contrôlée par un éminent groupe d'experts dont j'assurerai la présidence. Voilà donc une autre illustration des liens que nous sommes en train de bâtir, pas uniquement avec la collectivité universitaire, mais au sein de l'ONS et avec d'autres bureaux de statistique nationaux.

Les méthodes d'estimation seront complexes et recourront aux données de diverses sources :

- enquête de validation du recensement;
- analyses démographiques;
- dossiers administratifs.

Ces sources serviront à estimer le sous-dénombrement net à l'échelon régional et à modéliser des estimations régionales pour les niveaux géographiques inférieurs (p. ex., par l'application d'approches articulées sur la régression). Ensuite, si la chose est réalisable, on produira une base de microdonnées entièrement rajustée, ce qui débouchera sur des produits cohérents.

Enfin, on élaborera une méthodologie acceptable qui sera testée lors de la grande répétition de 1999.

10. CONCLUSION

J'espère vous avoir éclairé quelque peu sur l'approche que nous avons adoptée en vue du recensement de 2001, sur les changements envisagés et sur les innovations qui nous intéressent. L'essai de 1997 aura lieu en juin prochain et jouera un rôle déterminant dans le processus décisionnaire. Il nous permettra de choisir les questions primordiales avant la grande répétition de 1999 et de rédiger le livre blanc destiné au Parlement. Nous oeuvrons en étroite collaboration avec maints utilisateurs des données du recensement et avec la collectivité universitaire afin de créer ce qui deviendra, j'espère, un produit de qualité totale.

11. BIBLIOGRAPHIE

Census Division, *The One Number Census: Roma Chappell* (non publié). Office for National Statistics.

Clark, A., (1995), *Problems in large scale data collection : The 1996 Census*, Office for National Statistics.

Clark, A., (1995), *OPCS Census Development Programme*, Census Office for National Statistics.

Craig, J., (1995), *The Background – 1991 Census of England and Wales*, Census Division OPCS (now ONS).

Dugmore, K., (1995), *What do Users want from the 2001 Census?*, CACI.

Heady, P., (1995), *Census Validation Methods, with special reference to identifying which dwellings are occupied and to capture/recapture estimation*, Office for National Statistics.

Thomas, J., et Teague, A., (1996), *Neural Networks as a possible means for imputing missing Census data in the 2001 UK Census of Population*, Census Division, Office for National Statistics.

NORMES ET LIGNES DIRECTRICES CONCERNANT L'ERREUR NON DUE À L'ÉCHANTILLONNAGE: FOURNIR UN MEILLEUR SERVICE À MOINDRE PRIX

Richard D. Burgess¹

RÉSUMÉ

Fournir des données «prêtes à l'emploi» est l'un des objectifs principaux d'un organisme statistique. Ce dernier ne peut réussir sur un marché concurrentiel que si les utilisateurs sont convaincus que les données répondent à leurs exigences en matière d'information et de qualité. En règle générale, les utilisateurs doivent compter sur des données exactes, à jour et conformes à leurs propres besoins conceptuels, sur des données dont le prix est concurrentiel et qui se traduisent de plus en plus par des renseignements fondés sur des données longitudinales ou sur des données intégrées provenant de plusieurs sources. Les normes et les lignes directrices peuvent être utilisées judicieusement pour tenir compte des nombreuses exigences des utilisateurs. Qu'on les applique explicitement ou implicitement, en particulier pour résoudre la question de l'erreur non due à l'échantillonnage, les normes et les lignes directrices permettent d'accroître la productivité et la qualité, donc de fournir aux utilisateurs un meilleur service à moindre coût. L'auteur présente ici certaines idées sur le caractère opportun des normes et des lignes directrices dans le contexte de l'erreur non due à l'échantillonnage. Il décrit les points forts et les faiblesses d'un tel travail et donne des exemples d'applications à Statistique Canada.

MOTS CLÉS: Normes et lignes directrices; erreur non due à l'échantillonnage; service à la clientèle.

1. INTRODUCTION

Les clients d'un organisme statistique recherchent à coup sûr des données «prêtes à l'emploi». Il n'est pas rare de supposer que cela signifie que les utilisateurs doivent obtenir des données à la fois exactes et à jour, qui soient conformes à leurs besoins conceptuels ou analytiques, et dont le prix est concurrentiel.

Emprisonnés dans leurs exigences plus générales, les utilisateurs pourraient, non sans raison, s'attendre à ce qu'un bureau ou un organisme statistique, tout comme un organisme scientifique ou professionnel, applique des normes explicites en matière de concepts, de terminologie, de méthodes et de qualité. Ils pourraient également s'attendre à ce que les différences ou les écarts par rapport aux méthodes «types» aient un but démontrable et important. Plutôt que de s'y attendre, ils voudraient peut-être que ces normes soient fondées sur les connaissances les plus à jour, qu'elles satisfassent à leurs besoins actuels et soient constamment prises en compte dans les produits statistiques. Vu les progrès de la technologie, les utilisateurs s'attendront vraisemblablement à ce qu'une telle «normalisation» permette à un organisme efficace de créer une vaste gamme de produits statistiques transversaux, longitudinaux et intégrés et qu'elle satisfasse à leurs besoins de rapidité des systèmes micro-informatiques et d'économie.

Pour satisfaire les besoins et les attentes des utilisateurs, les normes fondamentales d'un organisme statistique doivent porter sur les connaissances et l'expérience de ses employés. D'autres normes et lignes directrices jouent un rôle dans la mise en place d'une gamme complexe de données prêtes à l'emploi et dans l'efficacité de leur diffusion: des normes et des lignes directrices servant à contrôler, à mesurer et à décrire l'erreur non due à l'échantillonnage dans les données statistiques.

¹Richard D. Burgess, Division des méthodes d'enquêtes sociales, 15-F, Immeuble R.H. Coats, Parc Tunney, Statistique Canada, Ottawa (Ontario), Canada K1A 0T6.

2. NATURE ET JUSTIFICATION DES NORMES ET DES LIGNES DIRECTRICES

2.1 Que sont les normes et les lignes directrices?

Les normes pourraient prendre la forme de règles ou de règlements visant à assurer la conformité ou l'uniformité. Pour l'erreur non due à l'échantillonnage et les organismes statistiques, l'uniformité ne constitue pas le contexte convenant à l'examen des normes. Les normes constituent plutôt une exigence de base, le modèle ou la méthode généralement acceptés, ou un jalon ou un modèle d'excellence. La plupart des normes comportent un élément de chacun de ces aspects.

On peut définir de la même façon les lignes directrices, mais il faut réduire le degré de conformité et, généralement, la portée des particularités. La souplesse de la mise en oeuvre peut être plus importante, réduisant du même coup tout élément d'uniformité. Du point de vue terminologique, les lignes directrices peuvent être plus facilement acceptables, mais peuvent être appliquées de façon moins constante.

2.2 Pourquoi la connaissance ne suffit-elle pas?

Si les exigences d'échantillonnage appropriées sont établies, l'erreur non due à l'échantillonnage constitue le pivot de la gestion de la qualité d'une enquête, ou de tout autre projet statistique, et de son coût. L'erreur non due à l'échantillonnage est contrôlée grâce à l'application de méthodes, mais pas nécessairement avec une notion claire de leurs coûts ou de leurs avantages globaux. Le mode de gestion de la qualité repose sur le discernement et non seulement sur les connaissances des nombreux participants au projet. Les personnes qui prennent des décisions courantes peuvent avoir un point de vue limité et peuvent être davantage préoccupées par la réalité des activités et de la gestion des coûts.

L'établissement de normes ou d'objectifs de qualité dès le départ permet d'orienter les connaissances là où il le faut et de les exprimer dans des critères et dans des méthodes acceptables. Ainsi, tous les intervenants, à partir de la conception, connaissent les éléments requis et peut-être ceux qui suffisent. Ces critères et pratiques peuvent ensuite tenir compte de la perspective globale, de même que de celle des programmes particuliers et de leurs utilisateurs. Par exemple, ces critères peuvent être insérés dans le contrôle de la qualité, dans les règles de révision, dans les exigences de suivi, dans des taux de réponse minimaux visés et dans la durée maximale d'une interview. Les pratiques peuvent prendre la forme d'exigences de contrôle de qualité, de vérification, d'essai de questionnaire et d'activités et processus, ou de communication des résultats et des méthodes, et de la qualité des données.

2.3 Quelles pourraient être les normes et les lignes directrices adéquates?

Voici certains arguments invoqués contre les normes:

- elles répriment la créativité et l'innovation, et freinent les gains de qualité et de productivité;
- elles se contentent du plus petit dénominateur commun et, tôt ou tard, ramèneront tous les éléments à ce niveau;
- l'effort nécessaire pour que les normes couvrent toutes les situations est beaucoup trop grand; par surcroît, les normes deviennent compliquées, inintelligibles ou simplistes, et elles ne sont pas mises en oeuvre.

Les chargés d'enquête et les intervenants semblent préférer les pratiques et les critères qu'ils choisissent eux-mêmes. Un tel choix se traduit trop souvent par une diversité inutile et, dans certains cas, par une baisse implicite de la qualité, par l'utilisation de méthodes, systèmes et mécanismes périmés, et par une réduction de la satisfaction des clients.

Les normes et lignes directrices ne doivent pas être appliquées de manière à limiter l'utilisation des connaissances et de l'expérience. Elles doivent être conçues pour accroître l'efficacité, l'efficacités ou la productivité. L'assurance de la qualité et les normes ne doivent pas se limiter à surveiller ou mesurer la qualité des données. Elles doivent avoir certains effets positifs à court et à long termes. Même si elles nécessitent une base solide de connaissances et de capacités, à certains égards les normes et lignes directrices appliquées à l'erreur non due à l'échantillonnage doivent être à l'avant-garde et être étayées d'activités de recherche-développement. Les normes doivent décourager

et éliminer le recours à des pratiques mauvaises ou de qualité inférieure, encourager les bonnes pratiques et fournir une orientation et établir des objectifs en matière de développement technique et d'amélioration du service à la clientèle.

Les normes peuvent naître du mariage des valeurs de l'organisation en matière de qualité et de la sensibilisation aux particularités des répondants. Elles doivent être fondées sur des principes bien articulés et faciliter l'application de ces derniers à des situations non prévues explicitement ou non envisagées. Elles doivent tenir compte de la mesure dans laquelle les programmes de l'organisation sont interdépendants et liés, et qu'ils partagent la même clientèle, la même base financière, la même renommée, et les mêmes connaissances et outils. Par ailleurs, il n'est pas nécessaire de produire toutes les normes à l'échelle de l'organisation. Les programmes n'ont pas tous la même importance, pas plus que des ressources équivalentes ou les mêmes exigences en matière de qualité. Chaque programme devrait posséder ses propres normes de qualité et méthodes d'assurance de la qualité.

3. UTILISATION DES NORMES ET DES LIGNES DIRECTRICES À STATISTIQUE CANADA

Au cours de la dernière décennie, des progrès technologiques importants - au niveau des opérations, de la statistique et des méthodes de calcul - ont permis au Bureau d'accroître l'efficacité de ses programmes et d'améliorer la qualité de leurs données. Cette période a également été marquée par des compressions budgétaires, un accroissement de la demande de données, tant au plan de la qualité que de la variété, et par une insistance plus marquée sur la qualité du service à la clientèle. Le Bureau crée une plus grande variété de données et de produits. Il a davantage recours à la collecte et à la diffusion électroniques des données, à leur intégration, à la création de données longitudinales, à l'analyse des données et au soutien de l'analyse des données portant sur des questions et politiques sociales et économiques. On demande au Bureau d'en faire plus avec moins, de se concentrer davantage sur les travaux complexes qui exposent aux répercussions de l'erreur non due à l'échantillonnage.

L'assurance et la gestion de la qualité de tous les programmes statistiques du Bureau doivent s'inscrire dans un cadre de pratiques, normes et lignes directrices, politiques et connaissances techniques établies. Comme dans tout autre organisme statistique, les connaissances et la formation constituent les normes et lignes directrices les plus importantes et les plus prépondérantes qui orientent les activités professionnelles des employés. Le recrutement, la formation et le perfectionnement du personnel, les comités internes et les comités consultatifs externes favorisent la constance des bonnes pratiques dans l'ensemble des programmes statistiques. Par conséquent, il existe des normes et lignes directrices implicites aux fins de la gestion de la qualité et du contrôle de l'erreur non due à l'échantillonnage. Le suivi de la non-réponse, la révision en vue d'assurer de l'information uniforme et complète, le contrôle de la qualité des activités, l'imputation et l'ajustement des poids en vue de tenir compte des incohérences et de la non-réponse constituent des pratiques courantes. Il incombe toutefois à chaque programme d'établir les critères, pratiques et méthodes particuliers, de même que le seuil de qualité acceptable, et de les justifier selon les circonstances, les contraintes, les possibilités et les objectifs.

Parmi les politiques, normes et lignes directrices particulières en vigueur à Statistique Canada et qui influent sur la gestion de l'erreur non due à l'échantillonnage, mentionnons:

- Lignes directrices pour la désaisonnalisation des données actuelles et révisées
- Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie
- Normes et lignes directrices concernant la documentation sur la qualité des données et la méthodologie
- Politique sur l'élaboration, l'essai et l'évaluation des questionnaires
- Lignes directrices pour l'essai et l'évaluation des questionnaires
- Normes et lignes directrices pour la présentation des tableaux dans les publications statistiques
- Normes et lignes directrices concernant la déclaration des non-réponses
- Avant-projet de politique sur les critères de qualité pour la diffusion de renseignements statistiques
- Normes et lignes directrices régissant l'application des critères de qualité pour la diffusion de renseignements statistiques
- Politique sur les estimations ayant des dates de référence future
- Lignes directrices concernant la qualité (un guide pour la planification en vue de fournir

des données de très grande qualité et d'évaluer la qualité)

La mise en oeuvre de ces politiques, normes et lignes directrices, etc., est revue et révisée à intervalles, habituellement par le Comité des méthodes et des normes, par le Comité consultatif des méthodes statistiques et par le Comité de vérification interne de Statistique Canada.

Les systèmes statistiques généraux mis au point par le Bureau pour l'ensemble de ses services constituent une norme, car ils permettent d'appliquer des pratiques et des méthodes uniformes et généralement à jour aux fonctions statistiques fondamentales. Ils sont habituellement utilisés pour les enquêtes nouvelles ou remaniées. Leur mise en oeuvre est relativement peu coûteuse.

Des normes sont également appliquées à la définition des unités géographiques et pour les classifications des secteurs d'activité et des professions. Nous poursuivons l'harmonisation et l'uniformisation des définitions appliquées aux variables par domaine, de même que l'élaboration de méta-bases de données. Ces éléments permettent ou permettront d'utiliser plus efficacement les données du Bureau qui déboucheront sur des renseignements plus complets à l'égard des concepts, des méthodes et de la qualité des données, et réduiront vraisemblablement le mauvais usage et l'interprétation erronée des données.

3.1 Résultat des normes?

Le présent document ne vise pas à donner un aperçu global des répercussions des politiques et des normes et lignes directrices connexes. Cependant, deux exemples importants pourraient justifier une certaine évaluation. Premièrement, la politique, de même que les normes et lignes directrices relatives à la conception des questionnaires ont été élaborées et sont faciles à mettre en oeuvre en raison de l'élaboration permanente de méthodes plus complexes et moins coûteuses d'essai des questionnaires. Bien qu'il subsiste des exemples de conception bien implantés et de ressources parfois limitées attribuées à la conception et à l'essai des questionnaires, la valeur de ces essais semble accrue. Cette situation est en grande partie attribuable aux résultats des essais et à leurs répercussions importantes sur les projets définitifs. Il s'agit là d'un bon exemple d'effort peu coûteux pour l'amélioration appréciable de la qualité des données et l'allègement du fardeau des répondants.

Une deuxième politique importante porte sur l'information des utilisateurs relativement à la qualité des données et à la méthodologie. Cette politique et les normes et lignes directrices connexes, qui en sont à leur troisième version, ont été mises en oeuvre pour la première fois à Statistique Canada en 1978. Les exigences de la politique constituaient d'abord un projet volontaire, un énoncé de valeurs, mais elles représentent maintenant des éléments obligatoires, de même que des lignes directrices complexes. L'exigence visant l'évaluation de la qualité constitue un principe de base de la politique, comme le point de vue selon lequel il convient d'offrir à l'utilisateur tous les renseignements publiés sur la qualité des données même si ces renseignements sont incomplets. Cette politique est généralement respectée à l'échelle des enquêtes ou des programmes statistiques (mais pas toujours à celle des produits) pour ce qui est de fournir des estimations de l'erreur d'échantillonnage. Les résultats dans le cas de l'erreur non due à l'échantillonnage sont moins satisfaisants. Les utilisateurs ont plus de difficulté à interpréter bon nombre de critères ou indicateurs de l'erreur non due à l'échantillonnage s'ils sont offerts sans restriction, dans le cas de données particulières ou même d'un tableau spécifique ou d'un ensemble de données. Il n'est pas rare que la seule évaluation efficace de la qualité relativement à l'erreur non due à l'échantillonnage réside dans les garanties fournies au moyen des descriptions de la méthodologie. Les garanties de qualité issues de mesures quantifiées, par exemple un taux de réponse, pourraient n'avoir qu'une signification relative. L'aise avec laquelle l'utilisateur se sert des données peut se fonder davantage sur la renommée du Bureau, sur le rendement antérieur d'un programme particulier et sur la description des méthodes qui font appel au bon sens, plutôt qu'à des résultats quantifiés ou historiques. Les mesures ne sont utiles que si l'on relève un problème général ou de grande envergure.

Malgré ces restrictions, il n'est pas déraisonnable que les programmes qui renferment de meilleures descriptions des méthodes et de la qualité sont ceux à l'égard desquels les paramètres de la qualité et de l'utilisation des mesures sont intégrés dans le processus de validation ou d'attestation des données. Il n'est donc pas surprenant de constater qu'il s'agit des programmes bien orientés et qui consacrent des ressources significatives à l'amélioration de la qualité. Ces programmes se défendent également bien sur la place publique pour ce qui est de la qualité, de l'interprétation et de la pertinence, facteur particulièrement important dans le cas des données relatives à des

questions touchant des opinions existantes et bien établies.

3.2 Normes de conception?

La politique concernant les critères touchant la qualité des données constitue un effort expérimental visant à éliminer certaines contraintes liées à la déclaration des paramètres d'évaluation de la qualité des données, à réduire les possibilités de commercialiser des données de piètre qualité et à exiger la spécification préalable des normes ou objectifs de conception pour assurer la qualité des données. La politique exige la classification de toutes les données dans l'une des catégories suivantes:

- qualité convenant pour les normes du projet
- qualité faible
- qualité inacceptable.

Les données pourraient être classées dans l'une des deux premières catégories d'après l'erreur d'échantillonnage ou dans l'une des trois catégories d'après l'erreur non due à l'échantillonnage. La classification doit se fonder sur une évaluation globale et elle doit être communiquée aux utilisateurs. Les données de qualité convenable doivent dominer les produits standard du Bureau. Ces produits ne doivent pas renfermer des données de qualité inacceptable. Les utilisateurs peuvent demander de recevoir des données, mais ils doivent être avisés de la qualité avant que l'une des parties ne prennent un engagement. Cette politique en est au stade expérimental dans le cadre de certains programmes statistiques de grande envergure.

On pourrait soutenir que cette politique dépasse les connaissances actuelles. En fait, elle oblige le fournisseur des données à faire ce qu'il exige des utilisateurs. Bien que l'établissement des catégories de qualité pourrait être discutable dans le cadre de toutes les utilisations éventuelles des données, il semble légitime dans le contexte des objectifs ou des attentes de la conception. Le plan d'assurance de la qualité précise clairement, sciemment ou non, le niveau de qualité suffisant, compte tenu des priorités, des objectifs et des contraintes.

La politique dépasse les pratiques actuelles pour ce qui est d'estimer ou d'attribuer l'erreur non due à l'échantillonnage à des cellules de données particulières. La classification et la détermination des cellules de données selon le niveau de l'erreur d'échantillonnage est conforme aux pratiques actuelles et courantes de déclaration de la qualité. Les pratiques actuelles d'estimation et de déclaration de l'erreur non due à l'échantillonnage n'attribuent toutefois l'erreur qu'à de vastes blocs de données. La viabilité à long terme de ce volet de la politique est étroitement liée à une évaluation à la fois meilleure et plus appropriée de l'erreur non due à l'échantillonnage. Ce volet s'appliquera plus particulièrement aux indicateurs des effets éventuels du redressement de la qualité, des effets de la pondération pour les non-répondants et des effets de l'imputation. Ces effets peuvent être évalués sans grande difficulté pour des cellules particulières. Il faudra tout de même interpréter l'impact éventuel sur la qualité du redressement, ou son absence, de chaque cellule. On pourrait examiner les effets, ou leur répartition, sur l'ensemble des programmes statistiques pour déterminer les éléments relativement tolérables ou atypiques. Il ne s'agit que d'un prolongement de l'utilisation actuelle des taux de réponse et d'imputation.

4. CONCLUSION

Les normes et les lignes directrices représentent simplement un moyen d'officialiser certains aspects des attentes de l'organisme à l'égard de chacun des éléments qui le compose. Elles ne produisent pas de gains de productivité ou une amélioration de la qualité, ni une réduction de l'erreur non due à l'échantillonnage. Elles sont plutôt un moyen de communication ou de renforcement des bonnes pratiques, et d'élimination des pratiques moins efficaces ou satisfaisantes. Leur mise en oeuvre peut se traduire par un coût considérable à l'égard des applications nouvelles et additionnelles. Cependant, compte tenu du fait que la pratique doit être appliquée, l'utilisation de normes et de lignes directrices pour orienter et assurer la mise en oeuvre peut réellement déboucher sur des gains d'efficacité grâce à la communication de méthodes efficaces et à l'expérience acquise par d'autres intervenants.

Les normes et les lignes directrices permettent d'intégrer les questions et principes de l'organisme dans les activités courantes. Si elles sont bien appliquées, elles entraîneront une amélioration de la qualité et son uniformisation et

pourraient déboucher sur la recherche de méthodes à la fois meilleures et moins coûteuses. Si elles ne permettent pas d'atteindre cet objectif, les normes et lignes directrices ne devraient pas exister.

5. BIBLIOGRAPHIE

Bradley, B. et Silins, J. (1995), Gestion des données pour le réseau d'information sur la santé du Canada: création d'un entrepôt virtuel d'information grâce à l'établissement de normes, de liens de coopération et de partenariats, Recueil du Symposium 1995 de Statistique Canada: Des données à l'information - méthodes et systèmes, Statistique Canada.

Burgess, R.D. (1990), Examen des critères de Statistique Canada relatifs des données diffusées, Recueil du Symposium 1990 de Statistique Canada: Mesure et amélioration de la qualité des données, Statistique Canada.

Early, J.F. (1990), La gestion de la qualité dans les programmes statistiques nationaux, Recueil du Symposium 1990 de Statistique Canada: Mesure et amélioration de la qualité des données, Statistique Canada.

Fellegi, I.P. (1995), Caractéristiques d'un système statistique efficace, Présentation à Washington Statistical Society, Washington D.C., 25 octobre 1995.

Kruskal, W. (1990), Introduction to Measurement Errors in Surveys, Rédacteur Biemer, Paul P., Groves, Robert M., Lyberg, Lars E., Mathiowetz, Nancy A., Sudman, Seymour, John Wiley & Sons, Inc. 1991.

Priest, G. (1995), Intégration des données: le point de vue de ceux qu'on relègue à l'arrière de l'autobus, Recueil du Symposium 1995 de Statistique Canada: Des données à l'information - méthodes et systèmes, Statistique Canada.

Woltman, H.F., et Thomas, K.F. (1990), Mesure de la qualité des données du recensement de 1990, Recueil du Symposium 1990 de Statistique Canada: Mesure et amélioration de la qualité des données, Statistique Canada.

SESSION 2

ERREURS LIÉES À LA BASE DE SONDAGE

MESURE DES ERREURS DU REGISTRE DES ENTREPRISES

N.Laniel, L.Mach, H.Finlay, S.Dionne¹

RÉSUMÉ

Le Registre des entreprises est une liste utilisée par Statistique Canada pour les enquêtes auprès des entreprises. Elle est établie en se fondant principalement sur deux fichiers administratifs indépendants. La première source est la liste des comptes de retenues sur la paye utilisés par les entreprises pour verser l'argent provenant des retenues à la source effectuées pour tous les employés. L'autre est le fichier des déclarations de revenus des entreprises aux fins de l'imposition. Le Registre des entreprises établi à partir de ces deux sources administratives est une base de sondage imparfaite, donc contenant divers types d'erreurs (c'est-à-dire sous-dénombrement, surdénombrement, classification erronée, taille inexacte et renseignements de contact incorrects). En premier lieu, on décrit les types d'erreur ainsi que leurs sources et leur effet sur les estimations d'enquête. Puis, on passe en revue et on analyse les diverses méthodes que l'on peut utiliser pour mesurer ces erreurs. On présente et examine aussi les résultats obtenus à l'aide de certaines de ces méthodes. Enfin, on examine brièvement les travaux qui seront poursuivis dans l'avenir en vue de mesurer le niveau des erreurs du Registre des entreprises.

MOTS CLÉS: Listes administratives; lacunes de couverture; classification erronée; renseignements de contact incorrects; méthodes de mesure des erreurs.

1. INTRODUCTION

Le Registre des entreprises de Statistique Canada est une banque centrale de données sur les entreprises du Canada basée sur deux listes administratives de Revenu Canada. Le Registre est la source principale des données utilisées pour définir les bases de sondage des enquêtes-entreprises annuelles et infra-annuelles du programme de la statistique économique. À ce titre, le Registre des entreprises (RE) a un rôle important à jouer dans la production d'un ensemble cohérent et intégré de statistiques économiques.

Les bureaux de statistique effectuent des enquêtes sur les entreprises pour produire les indicateurs économiques essentiels à l'analyse des politiques. Ces indicateurs doivent être fiables et refléter l'état actuel de la conjoncture économique. À l'endroit de la base de sondage, ce besoin de données fiables et à jour signifie que la liste d'entreprises utilisée pour une enquête, donc le RE, doit contenir des données précises, tenues aussi à jour que possible. Étant donné la nature dynamique de l'univers des entreprises, il ne s'agit pas là d'une tâche facile.

La précision et l'actualité des données du RE dépendent avant tout de deux facteurs, à savoir: (i) la qualité et l'actualité des deux listes administratives sources et (ii) la quantité de ressources consacrées au maintien de la base de données. Puisque les normes de qualité des données appliqués aux systèmes administratifs et statistiques diffèrent et que les ressources que l'on peut consacrer au maintien du RE sont limitées, il est inévitable que ce dernier contienne des erreurs. Celles-ci sont de plusieurs types, soit (i) les erreurs de couverture (par exemple, sous-dénombrement, unités non valides et doubles), (ii) les erreurs de classification (par exemple, classification erronée de l'activité économique ou géographique), (iii) les erreurs entachant les variables de stratification (par

¹ N. Laniel, L.Mach, H. Finlay et S. Dionne, Statistique Canada, 11^{ème} étage, Édifice R.H. Coats, Ottawa, Ontario, CANADA, K1A 0T6.

exemple, classe incorrecte pour le nombre d'employés ou le revenu) et (iv) les erreurs touchant les renseignements de contact (par exemple, adresse ou raison sociale incorrecte).

Il est important de mesurer la grandeur de chaque type d'erreur afin d'affecter efficacement les ressources limitées dont on dispose. Ces mesures permettent de réduire au minimum le niveau d'erreur et de produire des statistiques que les équipes d'enquête utilisent pour compenser les lacunes du RE, et ceci, au moment de l'établissement du plan de sondage ou du calcul des estimations.

Par exemple, dans le cas du plan de sondage, si on connaît la proportion d'unités non valides avant de tirer l'échantillon, on peut augmenter la taille de ce dernier en conséquence pour obtenir les variances d'échantillonnage souhaitées pour les estimations. Latouche et Hidiroglou (1987) ont mis au point une méthode de répartition d'échantillon qui tient compte de la proportion d'unités non valides non identifiées dans une base de sondage.

Un exemple d'utilisation de la mesure de l'erreur lors du calcul des estimations est celui de l'Enquête sur l'emploi, la rémunération et les heures de travail, où on se sert d'une mesure du sous-dénombrement entachant le RE pour corriger les estimations.

Si certaines erreurs entachant le RE peuvent être mesurées facilement, d'autres le sont difficilement. De surcroît, la mesure des erreurs est plus aisée dans le cas d'entreprises simples que dans celui d'entreprises dont la structure organisationnelle est complexe. Les méthodes de mesure des erreurs étant nombreuses, le présent article a pour objectif d'en passer quelques-unes en revue et de présenter les résultats obtenus ces dernières années en les appliquant au Registre canadien des entreprises.

La deuxième section de l'article traite de la conception et de la tenue à jour du RE et fournit les renseignements nécessaires pour définir les erreurs l'entachant et leurs sources. Celles-ci sont examinées à la section trois. À la section quatre, nous passons en revue les diverses méthodes applicables à la détermination du niveau d'erreur et présentons certains résultats obtenus en les appliquant au Registre canadien des entreprises. Enfin, à la section cinq, nous décrivons les travaux qui seront entrepris dans l'avenir quant à la mesure des erreurs entachant le RE.

2. CONCEPTION ET MAINTIEN DU REGISTRE DES ENTREPRISES

2.1 Conception

Le Registre des entreprises est une liste établie principalement avec deux sources de données administratives de Revenu Canada, plus précisément le fichier des déclarations annuelles de revenus des entreprises (T2) ainsi que des particuliers (T1) et le fichier des comptes de retenues sur la paye (Colledge, 1987). Les comptes de retenues sur la paye sont utilisés par les employeurs pour verser à Revenu Canada, ordinairement chaque mois, l'argent des cotisations au régime de pensions et au régime d'assurance-chômage, ainsi que d'autres retenues salariales. Or, il n'existe aucun identificateur commun permettant de coupler les données du fichier des déclarations de revenus et celles du fichier des comptes de retenues sur la paye. Donc, pour des raisons de temps et d'argent, le couplage des données de ces deux sources n'est pas faisable, étant donné le grand nombre d'entreprises (plus de deux millions) au Canada. Par conséquent, on a décidé de créer un registre comprenant deux composantes, l'une regroupant les entreprises grandes et (ou) complexes, où les fichiers de déclarations de revenus et de comptes de retenues sur la paye sont couplés, et l'autre consacrées aux petites entreprises à structure simple, ne contenant que les données du fichier des comptes de retenues sur la paye. Pour plus de renseignements sur la mise en oeuvre du système du RE, consulter Cuthill (1996).

La première composante, appelée partie intégrée (PI), regroupe les entreprises dont la structure organisationnelle est complexe et dont les activités s'étendent à plus d'une industrie ou province (environ 10 000 entreprises) et les grandes entreprises à structure simple (environ 88 000 entreprises) dont le revenu dépasse certains seuils. Ces seuils sont définis dans le Registre d'après le code à 2 chiffres de la Classification type des industries (CTI) et la province. Les entreprises groupées dans la PI sont à l'origine de plus de 70% du revenu total de chaque classe à 2 chiffres de la CTI par province. Des renseignements sur la structure des organisations complexes sont stockés

dans le RE pour que l'on puisse déterminer plus facilement auprès de quel élément de l'organisation il faut collecter les données. Toutefois, comme les deux sources de données administratives utilisées ne contiennent pas de renseignements organisationnels, on doit établir le profil des entreprises en communiquant directement avec celles-ci.

Le deuxième ensemble d'unités, appelé partie non intégrée (PNI), comprend les 800 000 entreprises employeuses simples et plus petites. Cette liste de petites entreprises est basée uniquement sur le fichier des comptes de retenues sur la paye. Ces comptes constituent une source de renseignements actuels pour la mise à jour du RE. Par exemple, un employeur peut les ouvrir ou les fermer n'importe quand. Leur inconvénient tient au fait que les entreprises sans employés ne sont pas couvertes. Pour prédire le nombre d'employés et le revenu, on se sert d'un modèle basé sur les remises actuelles des employeurs.

Ensemble, la PI et la PNI excluent 1 300 000 comptes de retenues sur la paye qui, pour la plupart, sont hors du champ d'observation (HCO) des enquêtes-entreprises. Toutefois, un petit nombre d'entre eux font partie du champ d'observation. Les unités exclues font partie de ce que l'on appelle la liste ZIP. Habituellement, les comptes de retenues sur la paye HCO sont les comptes dans lesquels aucun argent n'a été versé les douze derniers mois, les comptes-ménages, les comptes étrangers, les comptes ouverts par des entreprises qui ont cessé leurs activités économiques, les comptes de programmes gouvernementaux de travail spécial, les comptes de régime de pensions ou les comptes utilisés pour régler une succession. Les comptes de retenues sur la paye tombant dans le champ d'observation qui figurent sur la liste ZIP sont essentiellement des comptes ouverts par des entreprises canadiennes exploitées activement pour lesquelles on ne connaît pas le code approprié de la CTI ou de la Classification géographique type (CGT). On se réfère à ces comptes par l'appellation: unités non classifiées.

2.2 Maintien

Deux sources distinctes de données servent à maintenir le RE aussi à jour que possible, à savoir les données administratives de Revenu Canada et les enquêtes du RE destinées à collecter les données de base de sondage. Dans les paragraphes suivants, nous décrivons les diverses sources de données de mise à jour du RE ainsi que les mises à jour qu'elles permettent d'effectuer.

2.2.1 Données administratives de Revenu Canada

On distingue deux grandes sources de données de mise à jour axées sur les comptes de retenues sur la paye de Revenu Canada: (i) le fichier PAYDAC pour les ajouts et les suppressions d'entreprises, ainsi que les changements de raison sociale et d'adresse, et (ii) les formulaires PD20 pour les données de classification initiale.

Le fichier PAYDAC est une liste exhaustive de tous les comptes de retenues sur la paye ouverts par les employeurs. Il est mis à jour quotidiennement par Revenu Canada. Autrement dit, aussitôt qu'un employeur demande l'ouverture d'un compte, un nouvel enregistrement est ajouté au fichier. En outre, quand un employeur cesse ses activités économiques et règle tous les soldes exigibles, le compte est fermé et l'enregistrement correspondant est supprimé du fichier PAYDAC en janvier, au moins 12 mois après la fermeture du compte, à condition que celui-ci n'ait pas été rétabli.

Une fois par mois, Revenu Canada transmet à Statistique Canada la version la plus récente du fichier PAYDAC. Puis, on fait un exercice de couplage entre ce nouveau fichier et l'information se trouvant dans la PI du Registre. Les comptes du nouveau fichier qui sont appariés aux grandes unités sont assignés à cet ensemble et sont utilisés pour signaler les changements d'ordre juridique, opérationnel ou comptable de la structure de ces entreprises, ainsi que la création ou la dissolution éventuelle d'unités dans ces structures. L'indication de changements déclenche une prise de contact avec les grandes entreprises sous forme d'enquêtes du Registre des entreprises (voir la section 2.2.2).

Dans le cas des petites entreprises, l'utilisation des données du fichier PAYDAC est plus étendue. On s'en sert non seulement comme signaux de changement, mais aussi comme source principale de renseignements pour la tenue à jour de la liste des entreprises et des données de base de sondage. Les nouveaux comptes qui ne s'apparient pas à des unités de l'ensemble de grandes entreprises sont assignés à l'une de trois listes, selon le cas. Ils sont assignés

à la liste des petites entreprises, à titre de nouvelles entreprises, si des versements ont été effectués au cours des douze derniers mois et qu'on connaît tous les renseignements de classification. Ils sont inclus dans l'ensemble des comptes non classifiés si des versements ont été effectués au cours des douze derniers mois, mais qu'on ne connaît pas tous les renseignements de classification. Enfin, ils sont inclus dans l'ensemble d'unités HCO si aucun versement n'a été effectué au cours des douze derniers mois ou que l'unité répond à des critères d'exclusion.

En principe, chaque mois, les employeurs font à Revenu Canada des versements qui sont crédités à leurs comptes. Les renseignements concernant ces transactions sont enregistrés dans le fichier PAYDAC transmis à Statistique Canada mensuellement. Dans le cas des petites entreprises, les données sur les versements servent à mettre à jour les prédictions du nombre d'employés et du revenu brut de l'entreprise, au moyen de modèles, quand l'écart entre l'ancienne et la nouvelle valeur dépasse un seuil de tolérance. Aux entreprises dont le compte de retenues sur la paye a été inactif douze mois, on impute un revenu brut d'entreprise nul. Toutes les entreprises de la PNI auxquelles on impute un revenu brut d'entreprise nul sont supprimées de la liste et transférées à la liste ZIP (HCO).

Chaque fois qu'un employeur fait un versement, Revenu Canada lui envoie un reçu PD7AR. Une partie du formulaire PD7AR peut être détachée et renvoyée pour communiquer tout changement de raison sociale ou d'adresse. Les corrections sont intégrées au fichier PAYDAC, puis utilisées par Statistique Canada comme indication de changement dans le cas des grandes entreprises, et comme mise à jour directe dans le cas des petites.

Quand un employeur demande l'ouverture d'un compte de retenues sur la paye, Revenu Canada lui envoie un formulaire PD20 à remplir. Ce formulaire sert essentiellement à obtenir des données juridiques et opérationnelles (dont le nombre prévu d'employés et une description des activités de l'entreprise). Le temps pris pour renvoyer le formulaire dûment rempli à Revenu Canada peut varier de quelques jours à plusieurs mois.

Chaque semaine, Revenu Canada envoie une copie des formulaires PD20 reçus à Statistique Canada qui saisit les données qu'ils contiennent et les utilise pour initialiser les données de base de sondage concernant les petites entreprises. Ces formulaires représentent la principale source de codes initiaux de la CTI et de la CGT pour la plupart des nouvelles entreprises. Une fois que Statistique Canada a reçu la copie du formulaire PD20 d'une entreprise et saisi les données, l'entreprise est ajoutée à la composante non intégrée du Registre des entreprises, si les données de classification sont suffisantes et que l'entreprise effectue des versements. Sinon, elle est ajoutée à la liste ZIP des entreprises HCO si aucun versement n'est effectué, ou à la liste ZIP des entreprises non classifiées si des versements ont été effectués.

Dans le cas des grandes entreprises, les déclarations annuelles de revenus sont une autre source de données de mise à jour. Chaque année, Revenu Canada transmet une copie du fichier des déclarations de revenu à Statistique Canada. Ces déclarations indiquent les changements éventuels de la structure de l'entreprise qui n'ont pas été repérés grâce aux compte de retenues sur la paye et permettent de mettre à jour les données sur le revenu.

2.2.2 Enquêtes du Registre des entreprises

Statistique Canada effectue des enquêtes pour obtenir des données de base de sondage sur les entreprises pour lesquelles des éléments de données manquent, pour confirmer un changement décelé ou pour mettre les données du Registre à jour. Pour une discussion de certains points dont il faut tenir compte afin de décider des enquêtes du RE qu'il convient d'effectuer pour la mise à jour, consulter Colledge, Estevao et Foy (1987).

Dans le cas des petites entreprises, on se sert notamment du Rapport d'activité de l'entreprise. Il s'agit d'une enquête postale auprès des employeurs qui effectuent des versements, mais pour lesquels les données du formulaire PD-20 n'ont pas été transmises au RE dans les 90 jours après l'ouverture d'un compte ou pour lesquels les activités économiques décrites sur le formulaire PD20 n'ont pu être codées convenablement selon la CTI. Si le RE n'obtient pas de réponse de l'entreprise titulaire du compte de retenues sur la paye au moyen du Rapport d'activité de l'entreprise, un suivi est effectué grâce à l'Enquête de classification qui est effectuée par téléphone par le personnel des bureaux régionaux. L'Enquête de classification fournit des données supplémentaires, notamment des renseignements juridiques, des renseignements de contact et des renseignements administratifs.

L'Enquête de promotion est une autre enquête que l'équipe du RE effectue à l'heure actuelle auprès des petites

entreprises. Elle consiste essentiellement à communiquer avec toutes les petites entreprises dont le revenu brut dépasse le seuil de démarcation PNI/PI afin de confirmer que leur revenu est effectivement important. Le cas échéant, l'entreprise est promue à la composante des grandes entreprises et aucune donnée de base de sondage n'est collectée. Au lieu de cela, on établit un profil par réaction, dont nous parlerons plus bas, en vue de collecter les données. La promotion d'une unité à l'ensemble des grandes entreprises entraîne la suppression d'une petite unité et l'addition d'une grande unité dans le RE. Par contre, si la taille du revenu est jugée faible, l'entreprise demeure dans l'ensemble des petites entreprises et on collecte immédiatement les mêmes données de base de sondage que dans le cas de l'Enquête de classification.

L'Enquête sur les unités nouvellement enquêtées (ENE), effectuée tant auprès des grandes que des petites entreprises inscrites sur le Registre, visent les unités nouvellement sélectionnées dans l'échantillon d'une enquête économique. Elle permet de collecter les mêmes données de base de sondage que l'Enquête de classification. L'ENE est la source principale de données d'enquête destinées à compenser, grâce à l'estimation par domaine, les codes incorrects de la CTI figurant sur le RE.

L'Enquête de rétroaction de base de sondage est une autre enquête du RE effectuée auprès des grandes et des petites entreprises visées par les enquêtes économiques. Elle est effectuée auprès de toute unité pour laquelle une enquête économique indique un écart entre les données de la base de sondage et les renseignements fournis par le répondant au moment de la collecte de données économiques. Cette enquête du RE fournit les mêmes données de base de sondage que l'Enquête de classification.

En plus, on procède à l'établissement de profils. Cet exercice consiste à effectuer une enquête auprès des grandes entreprises en vue de collecter des données sur leurs structures opérationnelle, juridique et comptable, lesquelles sont souvent complexes. Farrall et Demmons (1987) examinent les problèmes que pose l'établissement des profils d'entreprise. Ces profils sont de deux types, à savoir le profil par réaction et le profil cyclique. Le profil par réaction correspond, par exemple, à un miniprofil établi consécutivement à la modification des données de l'ensemble de comptes de retenues sur la paye d'une grande entreprise ou des données de la déclaration de revenus. Clark et Lussier (1987) décrivent en détail l'utilisation des données administratives pour établir les profils par réaction. L'établissement du profil cyclique, quant à lui, est un exercice approfondi et périodique qui consiste à définir le profil des grandes entreprises dans le but de tenir les données sur ces entreprises aussi à jour que possible. À l'heure actuelle, la durée du cycle est de deux ans environ, mais on vise à la réduire à un an à mesure qu'on gagnera de l'expérience et qu'on améliorera les instruments utilisés.

3. DÉFINITION ET SOURCES DES ERREURS

Il est généralement admis qu'il est impossible de créer des registres d'entreprises parfaits, parce que les ressources que l'on peut consacrer à leur maintien sont limitées et qu'il est difficile de les tenir à jour (Tupek, Copeland et Waite, 1988). Nous allons maintenant définir les erreurs de base de sondage, en mentionner les causes et décrire leur effet sur les estimations calculées d'après les données des enquêtes économiques. Les erreurs décrites dans la présente section ne sont pas particulières au Registre des entreprises de Statistique Canada. Au contraire, elles entachent aussi les registres des autres pays, voir Konschnik (1988) en ce qui concerne les bases de sondage américaines.

3.1 Unités manquantes

Les unités manquantes sont les entreprises qui devraient figurer dans la PI ou la PNI du Registre des entreprises mais n'y figurent pas. Dans le cas du Registre canadien des entreprises, on distingue six catégories d'unités manquantes. La première correspond aux entreprises non classifiées de la liste ZIP qui effectuent des versements dans un compte de retenues sur la paye et avec lesquelles des contacts sont en train d'être établis via l'Enquête de classification. La deuxième catégorie comprend les unités classées erronément dans la catégorie des entreprises hors du champ d'observation (HCO) de la liste ZIP. Par exemple, les entreprises employeuses exploitées activement qui ont omis d'effectuer des versements à Revenu Canada pendant plus de 12 mois. La troisième comprend les nouvelles entreprises employeuses exploitées activement qui ont ouvert récemment un compte de

retenues sur la paye, mais qui ne sont pas encore inscrites sur le RE. La marche à suivre pour ajouter de nouveaux comptes actifs cause un retard d'enregistrement de deux à six semaines. La quatrième catégorie inclut toutes les petites entreprises non employeuses exploitées activement qui sont exclues du RE par conception de ce dernier, la PNI est fondée uniquement sur les données administratives tirées des comptes de retenues sur la paye. La cinquième regroupe les grandes entreprises non employeuses exploitées activement qui ne sont pas encore inscrites dans le RE à cause de retards dus au traitement des données administratives tirées des déclarations annuelles de revenus. Enfin, la sixième catégorie comprend les sous-unités de grandes organisations déjà représentées dans la PI qui n'ont pas encore été ajoutées au RE en raison du temps que demandent l'établissement du profil d'unités complexes et le traitement des données résultantes.

Puisqu'elles ne peuvent être incluses dans les bases de sondage des enquêtes, les unités manquantes introduisent un biais négatif dans les estimations.

3.2 Unités non valides

Les unités non valides sont celles qui ne devraient figurer ni dans la composante intégrée ni dans la composante non intégrée du RE. Elles se classent en trois catégories. La première comprend les titulaires de comptes de retenues sur la paye de la PNI qui effectuent des versements mais ne participent à aucune activité économique (par exemple, comptes de ménage). La deuxième catégorie comprend les entreprises employeuses inactives de la PNI qui ont effectué des versements au cours des 12 derniers mois. Soulignons qu'il s'agit d'une lacune inhérente à la conception du registre, car on s'efforce de réduire au minimum le dénombrement dû aux entreprises employeuses exploitées activement qui tardent à effectuer leurs versements à Revenu Canada. Enfin, la troisième catégorie regroupe les grandes entreprises non employeuses inactives, ou des parties de celles-ci, qui n'ont pas encore été reconnues comme non exploitées activement grâce aux données des déclarations annuelles de revenus, des profils par réaction ou des profils cycliques.

Aucune source de renseignements ne permet de repérer toutes les unités non valides. Par conséquent, ces dernières peuvent causer un biais et une augmentation de la variance des estimations par sondage. Les estimations sont entachées d'un biais positif si les unités non valides ne sont pas reconnues comme telles au moment de l'enquête par sondage et qu'on impute des données. En revanche, si on repère les unités non valides au moment de l'enquête par sondage, le biais est réduit, mais la variance augmente, puisque la taille effective de l'échantillon diminue.

3.3 Doubles

Les doubles correspondent à des unités qui sont représentées plus d'une fois sur le RE. Il existe deux catégories de doubles. La première regroupe les unités enregistrées à la fois dans la PI et dans la PNI en raison de l'imperfection du procédé de couplage du fichier PAYDAC à celui de la PI. Le couplage du fichier des nouveaux comptes de retenues sur la paye au fichier de la PI aboutit au non-appariement d'enregistrements qui sont, en fait, déjà représentés dans la PI. La deuxième catégorie comprend les unités de la PI qui sont enregistrées plus d'une fois en raison de l'appariement imparfait du fichier des comptes de retenues sur la paye et du fichier des déclarations de revenus.

Il convient de noter qu'en principe, les doubles ne posent pas de problème dans le cas de la PNI pour deux raisons. Premièrement, les comptes de retenues sur la paye qui forment la PNI possèdent des identificateurs uniques qui rendent impossible la présence de doubles parmi les comptes. Deuxièmement, pour effectuer les enquêtes, on sélectionne des comptes de retenues sur la paye dans le fichier de la PNI, puis on communique avec les entreprises titulaires de ces comptes afin de confirmer la liste de leurs comptes de retenues sur la paye. On se sert alors de cette information pour corriger les poids à l'étape du calcul des estimations. Cette méthode donne de bons résultats à condition que les entreprises fournissent la liste complète de leurs comptes de retenues sur la paye quand on la leur demande.

Les enregistrements qui figurent à la fois dans la PI et la PNI entachent les estimations d'enquête d'un biais positif, puisque la population observée est alors plus grande que la population cible.

3.4 Classification incorrecte

Une unité est mal classifiée si on lui attribue un code de la CTI ou de la CGT incorrect. Une erreur de classification peut survenir quand les données utilisées pour le codage sont inexactes ou que la personne qui effectue le codage commet une erreur. L'erreur de classification peut aussi venir du fait qu'on ignore que l'unité a changé de classe industrielle depuis la dernière collecte de données.

Les unités mal classifiées causent un biais de couverture et (ou) une augmentation de la variance des estimations d'enquête. Si certaines unités ne sont pas incluses dans la population que vise une enquête sur une industrie particulière à cause de leur classification erronée, les estimations sont entachées d'un biais négatif. Inversement, quand des unités sont incorrectement incluses dans la population observée en raison de leur classification erronée, les estimations sont entachées d'un biais positif si les unités mal classées ne peuvent être décelées lors de la collecte des données. Cependant, si la collecte permet de dépister les unités incorrectement classifiées, on traite celles-ci comme des entreprises inactives, ce qui élimine le biais positif, mais augmente la variance, car il est nécessaire de procéder à une estimation par domaine.

3.5 Mesure erronée de la taille

Par mesure erronée de la taille d'une unité, on entend une mesure qui s'écarte beaucoup de la valeur réelle. Dans le cas des grandes unités de la PI, ce type d'erreur survient parfois parce que la source principale de données utilisée pour la mise à jour de la taille de l'entreprise est la déclaration de revenus qui est périmée aussitôt qu'elle est produite. Dans le cas des petites unités, la mesure de la taille se fonde sur des modèles qui, par définition, introduisent un certain biais et une certaine variabilité des valeurs prévues.

L'erreur de mesure de la taille peut avoir une incidence importante sur les estimations par sondage. Au stade de l'établissement du plan de sondage, elle peut empêcher d'obtenir la stratification optimale selon la taille et la répartition optimale de l'échantillon. En outre, au stade du calcul des estimations, elle risque de produire de nombreuses valeurs aberrantes pour les unités dont la taille prédite est beaucoup plus faible que la taille réelle.

3.6 Renseignements de contact inexacts

Les renseignements de contact pour une entreprise sont inexacts quand ils empêchent de collecter les données à l'occasion d'une enquête économique. Cette situation se produit parfois parce que les renseignements de contact que l'on a tirés des données administratives de Revenu Canada et inscrits dans le RE ne conviennent pas toujours à la collecte de données d'enquête.

Des renseignements inexacts aboutissent parfois à une non-réponse quand la personne-ressource appropriée ne peut être rejointe en temps voulu. Or, la non-réponse peut biaiser les estimations par sondage et augmenter leur variance, qu'on effectue ou non une imputation.

L'inexactitude des renseignements de contact peut aussi entraîner des erreurs systématiques de réponse quand la structure de l'entreprise est complexe. En effet, dans ce cas, le fait de ne pas contacter la bonne personne peut avoir pour conséquence de ne pas obtenir toutes les données nécessaires sur l'organisation et de ne collecter des données que sur une partie celle-ci, c'est-à-dire celle que connaît la personne interrogée.

4. MESURE DES ERREURS

Essentiellement, les lacunes du RE qu'il convient d'évaluer sont les erreurs de couverture (doubles, unités non valides, unités manquantes et unités classifiées incorrectement de sorte qu'elles sont incluses dans la mauvaise population étudiée) et l'inexactitude de certaines données élémentaires de la base de sondage (codes de la CTI, codes de la CGT, mesure de la taille, raison sociale et adresse). Idéalement, on devrait exprimer les mesures en fonction du nombre d'unités, pour déterminer ce que coûterait la réduction du niveau d'un type donné d'erreur, ainsi qu'en fonction des mesures de taille pour l'incidence des erreurs sur les estimations par sondage. En outre, on

devrait effectuer les mesures périodiquement, pour pouvoir observer l'amélioration ou la détérioration de la qualité des données, donc, pour pouvoir prendre les mesures appropriées, en temps voulu, et corriger les estimations. Ce sont là les objectifs que s'est fixée la Division du Registre des entreprises de Statistique Canada en ce qui concerne l'évaluation de la qualité.

Dans la présente section, nous examinons les types d'erreur difficiles à mesurer, puis nous passons en revue les méthodes utilisées jusqu'à présent pour mesurer les erreurs qui entachent les registres des entreprises et, enfin, nous présentons et examinons certains résultats.

4.1 Erreurs difficiles à mesurer

À l'heure actuelle, le registre compte deux groupes d'unités pour lesquelles il est difficile d'obtenir des mesures d'erreur. Il s'agit des unités complexes, d'une part, et de celles inscrites erronément sur la liste des unités HCO, d'autre part.

a) Unités complexes

On ne dispose, à l'heure actuelle, d'aucune mesure des erreurs entachant les données sur les unités complexes de la PI du Registre des entreprises. On sait uniquement que, selon l'Enquête sur l'emploi, la rémunération et les heures de travail, le nombre déclaré d'employés par les entreprises pour lesquelles la Division du RE a établi un profil récemment a augmenté. Cette augmentation peut être considérée comme une mesure du sous-dénombrement des sous-unités des organisations complexes inscrites dans le Registre des entreprises.

Aucune mesure de la qualité des données n'a encore été effectuée pour deux raisons. Premièrement, le fardeau de réponse de ces entreprises est déjà lourd, sans même compter leur participation à l'établissement des profils, car elles sont sélectionnées avec certitude dans l'échantillon de nombreuses enquêtes. Deuxièmement, l'établissement actuel du profil de ces entreprises complexes est déjà un exercice coûteux. Par conséquent, il ne conviendrait pas de mettre au point un instrument de mesure susceptible d'augmenter le fardeau de réponse et les coûts d'enquête. La seule approche viable consisterait à se servir des résultats des profils par réaction et des profils cycliques actuels. Il serait souhaitable de poursuivre les travaux dans ce domaine.

b) Unités considérées erronément hors du champ d'observation

Pour estimer de façon non biaisée le nombre d'unités jugées erronément HCO, il faut échantillonner l'ensemble des unités HCO non appariées à une unité de la PI, puis étudier directement les unités échantillonnées (pour s'assurer qu'il s'agit d'entreprises employeuses exploitées activement et pour vérifier les codes de la CTI et de la CGT qui leur sont attribués). Malheureusement, il faut tirer un échantillon de grande taille pour obtenir des estimations fiables, puisque la proportion d'unités jugées erronément HOC est, en principe, très faible. De surcroît, il pourrait être difficile et onéreux d'essayer de communiquer avec les unités d'un échantillon composé principalement d'unités inactives, puis de déterminer si ces unités font ou non partie du champ d'observation. Il n'existe aucune méthode évidente de mesure de ce sous-dénombrement. On devrait envisager de s'appuyer sur des sources externes de renseignements.

4.2 Méthodes utilisées pour mesurer les erreurs

Nombre de méthodes peuvent être appliquées pour mesurer l'importance des erreurs qui entachent le RE. Nous passons en revue, aux paragraphes suivants, les quelques-unes utilisées jusqu'à présent.

a) Enquête de mesure de la qualité

Statistique Canada a effectué à plusieurs reprises une Enquête de mesure de la qualité (EMQ) pour évaluer la qualité des données des bases de sondage (voir Lorenz et Laniel, 1992). La dernière, menée en 1995, a donné lieu à la collecte de la plupart des données élémentaires du Registre des entreprises pour les grandes unités à structure simple ainsi que pour les petites unités. L'enquête a été effectuée par téléphone à partir des bureaux régionaux auprès d'un échantillon aléatoire stratifié de 5 000 unités et la collecte des données a duré 3 mois. Le taux de

réponse a été d'environ 90 %.

b) Source interne

Comme nous l'avons mentionné à la section 2, le Registre des entreprises contient une liste d'unités non classifiées que l'on peut consulter. On procède à l'heure actuelle à l'examen de toutes ces unités en se basant sur le formulaire de comptes de retenues sur la paye, sur le Rapport d'activité de l'entreprise ou sur l'Enquête de classification. Autrement dit, dans un avenir plus ou moins proche, toutes ces unités seront classifiées. Ici, la méthode de mesure du sous-dénombrement consiste à sélectionner une période de référence, à obtenir la liste de toutes les unités non classifiées dans la catégorie considérée et d'attendre un certain nombre de mois de façon à ce que la grande majorité soit classifiée. Une étude antérieure indique que le temps médian nécessaire pour classifier les unités non classifiées est de huit mois. Cette étude donne à penser qu'après un an, la plupart des unités non classifiées le seront finalement. Cette méthode produit des estimations biaisées si les unités ne sont pas toutes classifiées.

c) Sources externes

On peut s'appuyer sur trois sources externes de données pour estimer le sous-dénombrement du Registre des entreprises, à savoir l'Enquête sur l'emploi, la rémunération et les heures de travail, l'Enquête sur la population active et la Liste des déclarants de Revenu Canada. Le champ d'observation de l'Enquête sur l'emploi, la rémunération et les heures de travail inclut les comptes de retenues sur la paye non classifiés, donc fournit une estimation du nombre d'employés des entreprises titulaires de ces comptes. Cependant, cette estimation ne fournit pas de ventilation par industrie.

L'Enquête sur la population active fournit une estimation du nombre de travailleurs indépendants incluant ceux qui emploient des salariés dans leur entreprise. Cette estimation constitue la limite supérieure du nombre d'entreprises non employeuses, qui est aussi le nombre d'employés dans ces entreprises.

La Liste des déclarants de Revenu Canada fournit la liste des déclarants qui travaillent à leur propre compte incluant ceux qui emploient des salariés dans leur entreprise. On peut se servir de cette liste pour calculer la limite supérieure du nombre d'entreprises non employeuses et du revenu total de ces entreprises.

d) Étude par couplage des enregistrements

Il est possible d'estimer le nombre d'enregistrements qui figurent à la fois dans la composante intégrée et dans la composante non intégrée du Registre des entreprises (i.e. doubles) en couplant un échantillon d'unités de la PNI aux unités de la PI, puis en évaluant manuellement l'exactitude des appariements. Une étude de ce genre a été effectuée en 1992 (Chun et coll., 1993) auprès de 13 400 unités de la PNI sélectionnées par échantillonnage aléatoire stratifié.

e) Étude des données sur le revenu obtenues par modélisation

Dans le cas des entreprises de la PNI, on évalue annuellement les erreurs qui entachent les prédictions du revenu brut des entreprises au moyen de modèles, en comparant la valeur prévue à la valeur calculée d'après les données de l'Enquête mensuelle sur le commerce de détail. Cette étude consiste fondamentalement à calculer des corrélations (consulter Dionne et Hawley, 1996).

4.3 Résultats

Examinons maintenant les résultats de l'estimation des erreurs qui entachent le Registre des entreprises effectuée ces dernières années en appliquant les méthodes susmentionnées.

a) Unités manquantes

Le nombre d'entreprises non classifiées inscrites sur le Registre des entreprises représente environ 1,8 % de la population totale de ce dernier. D'après l'Enquête sur l'emploi, la rémunération et les heures de travail, le nombre

total d'employés de ces unités représente 0,6 % du nombre total d'employés des entreprises employeuses. Le fait que les unités non classifiées soient ordinairement de nouvelles entreprises de petite taille explique cette faible proportion. Bien qu'on ne possède aucun chiffre catégorique quant à la proportion du revenu total imputable aux entreprises non classifiées, on peut supposer que ces dernières produisent moins de 0,6 % des recettes totales des entreprises employeuses, étant donné que la répartition des revenus est plus asymétrique que celle du nombre d'employés.

Une étude des unités non classifiées effectuée en 1995 montre que la répartition de ces unités selon la branche d'activité est semblable à celle des entreprises classifiées sur le Registre des entreprises.

En regroupant les estimations du nombre de travailleurs indépendants calculées d'après les données de l'Enquête sur la population active de 1994 et la Liste des déclarants de Revenu Canada pour la même année, on estime que la limite supérieure de la proportion d'entreprises employeuses dans l'ensemble des entreprises est inférieure à 60 %. En ce qui concerne le nombre d'employés, l'estimation correspond à moins de 10 % du nombre d'employés dans toute la population. Enfin, pour ce qui est du revenu, on estime que les entreprises non employeuses produisent moins de 1 % du total.

b) Unités non valides

D'après l'Enquête de mesure de la qualité de 1995, la proportion d'unités non valides se chiffre à 14 % dans la population de grandes entreprises à structure simple de la PI du Registre et à 9,1 % dans la population de petites entreprises de la PNI. L'écart peut être attribué au fait que la méthode de mise à jour de la PI (c'est-à-dire l'établissement de profils) ne permet pas de dépister aussi rapidement les unités disparues que la méthode de mise à jour de la PNI (c'est-à-dire l'examen des comptes de retenues sur la paye dans lesquels aucun versement n'a été effectué depuis 12 mois). Il convient de souligner que les deux chiffres peuvent être entachés d'un biais négatif, puisqu'il est vraisemblable que la population d'entreprises non répondantes contienne une plus forte proportion d'unités non exploitées que d'unités exploitées activement.

c) Doubles

L'Enquête de mesure de la qualité effectuée en 1991 auprès d'un échantillon d'unités de la PNI (voir Lorenz et Laniel, 1992) indique que 0,2 % seulement des petites unités était un double d'une grande unité. Il convient de noter que le succès de la méthode suivie pour déceler les doubles dépendait de la volonté des entreprises de fournir la liste complète de leurs comptes de retenues sur la paye. Cette information étant considérée confidentielle par certaines, on peut supposer que le taux de double est entaché d'un biais négatif.

L'étude par couplage des enregistrements de 1992 indique qu'environ 1,6 % des unités de la partie non intégrée étaient des doubles des unités de la partie intégrée. Cette estimation du taux de doubles est supérieure à celle résultant de l'enquête de 1991. Cela semble confirmer que les répondants n'ont pas fourni aux intervieweurs les renseignements nécessaires pour le couplage. Par ailleurs, on soupçonne aussi qu'un certain nombre de faux appariements n'ont pas été repérés.

d) Classification erronée

L'EMQ de 1995 a également fourni les taux d'erreur de classification par industrie pour les unités exploitées activement du Registre des entreprises. Ces taux sont présentés dans les tableaux qui suivent.

Le tableau 1 contient les estimations du pourcentage d'unités classées dans une industrie incorrecte pour 4 niveaux de la Classification type des industries, selon la taille de l'entreprise (c'est-à-dire partie intégrée ou partie non intégrée). À noter qu'aux niveaux à 3 et à 4 chiffres, ces taux d'erreur devraient être interprétés avec prudence. En effet, pour nombre d'industries, aucun code à 3 ou à 4 chiffres n'est produit car seule la production d'un code d'industrie à 2 chiffres ou à 3 chiffres est requise.

D'après le tableau, le taux d'unités classées dans la mauvaise industrie est fort semblable pour les grandes et pour les petites entreprises, ce qui donne à penser que si on calculait la part du nombre total d'employés ou de revenu

correspondant à ces taux, les proportions seraient du même ordre.

Tableau 1. Estimations de la proportion d'unités classées dans une industrie incorrecte

Catégorie	PI simple	PNI
Division incorrecte (groupe à 2 chiffres)	7,4 %	7,5 %
Grand groupe incorrect (2 chiffres)	10,4 %	9,6 %
Groupe incorrect (3 chiffres)	15,9 %	16,5 %
Classe industrielle incorrecte (4 chiffres)	18,7 %	22,7 %

Le tableau indique que le taux moyen de sous-dénombrement d'une division se chiffre à 7,5 %. Autrement dit, une enquête dont la population cible appartient à une seule division pourrait produire une surestimation importante du nombre d'entreprises dans cette division si l'exactitude de l'activité industrielle des unités échantillonnées n'est pas confirmée. Il convient de souligner que ces taux pourraient être sous-estimés, puisque l'intervieweur demande au répondant de confirmer la description fournie par le Registre des entreprises (c'est-à-dire une méthode dépendante).

Cependant, demander au répondant de fournir une description de ses activités sans lui communiquer les données du Registre des entreprises pourrait produire une surestimation, car une nouvelle formulation de la description pourrait être interprétée à tort comme une activité différente.

L'EMQ de 1995 visait aussi à estimer la proportion d'unités manquantes dans chaque division. Selon les résultats, la proportion varie de 1 % à 15 %. Il convient de souligner que, pour une enquête sur une industrie particulière, cette source de sous-dénombrement est vraisemblablement la plus importante.

L'EMQ de 1995 fournit les taux d'erreur de classification géographique pour les unités exploitées activement du Registre des entreprises. Ces taux sont présentés au tableau 2 pour deux niveaux géographiques (c'est-à-dire la province et la région métropolitaine de recensement/agglomération de recensement) selon la taille de l'entreprise (c'est-à-dire PI ou PNI).

Tableau 2. Estimations du taux d'erreur de classification géographique

Catégorie	PI simple	PNI
Province incorrecte	0,05 %	0,16 %
RMR/AR incorrecte	3,2 %	2,2 %

On note que les erreurs sont très faibles au niveau provincial et faibles au niveau infraprovincial.

e) Mesure incorrecte de la taille

L'étude de 1995 visant à comparer le revenu brut d'entreprise prévu aux données de l'Enquête sur le commerce de détail a fourni les résultats que voici. Un premier calcul de corrélation directe a produit un coefficient de corrélation de 0,29, valeur qui est assez faible. Cependant, après qu'on ait éliminé 1 % des valeurs extrêmes du revenu brut d'entreprise prévu, la valeur du coefficient est passée à 0,61, ce qui indique une bonne corrélation. L'augmentation de la corrélation après l'élimination des valeurs extrêmes montre que la variance de l'erreur du modèle est relativement forte.

f) Renseignements de contact inexacts

L'EMQ de 1995 fournit les taux d'erreur relatifs aux renseignements de contact pour les unités exploitées activement. Ces taux sont présentés au tableau 3 pour la raison sociale et le code postal, selon la taille de l'entreprise (c'est-à-dire PI et PNI).

Tableau 3. Estimations des taux d'erreur relatifs aux renseignements de contact

Catégorie	PI simple	PNI
Raison sociale incorrecte	10,1 %	25,6 %
Code postal incorrect	21,6 %	11,4 %

En ce qui concerne la raison sociale, le taux d'erreur est plus élevé pour les entreprises de la partie non intégrée que pour celles de la partie intégrée. Cette situation reflète certainement le fait que les données de mise à jour de la partie intégrée sont tirées des déclarations de revenus (qui se rapportent à des personnes morales). Pour les unités de la partie non intégrée, l'EMQ de 1991 indique un taux d'erreur de 12 %. L'écart est dû à l'inclusion des fautes d'orthographe dans le calcul du taux d'erreur de 1995. Le taux d'erreur calculé pour le code postal reflète aussi l'utilisation de sources de données de mise à jour différentes. L'adresse de l'entreprise diffère souvent de l'adresse de la personne morale, mais se rapproche de l'adresse de l'entité opérationnelle. On notera qu'il est impossible de déterminer l'effet des erreurs entachant les renseignements sur les personnes-ressources d'après les taux susmentionnés. En fait, l'effet réel dépend de la méthode de collecte de données utilisée pour effectuer l'enquête.

5. CONCLUSION

Comme nous l'avons mentionné à la section 2, la conception et le maintien du Registre des entreprises sont des activités complexes, parce qu'elles s'appuient sur l'utilisation de sources de données administratives et sur des enquêtes pour modéliser le monde des affaires. Cette complexité entraîne les types d'erreur (erreurs de couverture, erreurs de classification et renseignements de contact erronés) et les sources d'erreur dont il est question à la section 3.

À la section 4, nous avons montré que la mesure du niveau de ces erreurs n'est pas une tâche simple, particulièrement dans le cas d'unités complexes ou d'unités manquantes. Nous avons montré également que certains taux d'erreurs sont importants. Par exemple, les erreurs de classification aux termes de la CTI constituent probablement une composante importante des biais et (ou) de l'erreur d'échantillonnage qui entachent les estimations par sondage. Au moment de l'examen de la répartition des ressources pour le maintien actuel du Registre des entreprises, on devrait envisager d'investir davantage dans le dépistage des sources d'erreur et dans l'amélioration des méthodes de codage selon la CTI. Il convient de souligner que, dans un avenir proche, la mise en oeuvre de la Classification type des industries de 1997, (également appelée Système de classification des industries de l'Amérique du Nord) donnera l'occasion d'améliorer la qualité des codes de la CTI. De juillet 1997 à décembre 1998, on communiquera avec à peu près la moitié des entreprises qui figurent dans le RE afin de confirmer leur branche d'activité.

Certaines améliorations pourraient être apportées aux mesures de la qualité du RE. Par exemple, on devrait mettre au point des méthodes de mesure des erreurs pour les unités complexes, afin de broser un tableau plus complet de la qualité du registre. On devrait aussi s'efforcer d'améliorer les méthodes de couplage visant à dépister les doubles ainsi que les méthodes d'évaluation des appariements décelés. De surcroît, on devrait obtenir les données économiques sur les doubles repérés pour évaluer l'importance de ceux-ci en ce qui a trait au revenu et à l'emploi. Enfin, on devrait modifier le plan de l'Enquête de mesure de la qualité pour permettre de bien estimer les erreurs en regard des mesures de taille des entreprises (c'est-à-dire le revenu et l'emploi). Cet exercice serait fructueux, car les estimations axées sur la taille pourraient servir à la correction des estimations tirées des enquêtes visant une industrie particulière.

6. BIBLIOGRAPHIE

Chun, B., Hutchinson, D. et Patak, Z. (1993). Methodology Report of IP/NIP Duplication on the Business Register. Document interne, Division des méthodes d'enquêtes-entreprises, Statistique Canada, Ottawa.

Clark, C. et Lussier, R. (1987). The Use of Administrative Data for Initial and Subsequent Profiles of Economic Entities. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Colledge, M., Estevao, V. et Foy, P. (1987). Experiences in the Coding and Sampling of Administrative Data. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Colledge, M. (1987). Utilisation des données administrative dans le contexte du projet de remaniement des enquêtes-entreprises, Recueil du symposium 1987 de Statistique Canada, Ottawa.

Cuthill, I. (1996). The Canadian Business Register. Document interne, Division du développement de système, Statistique Canada, Ottawa.

Dionne, S. et Hawley, M. (1996). Testing of the Final 1993 Gross Business Income Tables. Division des méthodes d'enquêtes-entreprises, rapport interne, Statistique Canada, Ottawa.

Farrall, K. et Demmons, P. (1987). Profiling for Statistics Canada's Central Frame Data Base. Business Survey Redesign Project Working Paper, Statistique Canada, Ottawa.

Konschnik, C.A. (1988). Coverage Error in Establishment Surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Latouche, M. et Hidioglou, M.A. (1987). Sample Size Determination and Allocation for the Monthly WRTS. Division des méthodes d'enquêtes-entreprises, rapport interne, Statistique Canada, Ottawa.

Lorenz, P. et Laniel, N. (1992). Measuring the Quality of the Business Register: Methodology and Results. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Tupek, A.R., Copeland, K.R. et Waite, P.J. (1988). Sample Design and Estimation Practices in Federal Establishment Surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 298-303

L'ESTIMATION DE L'ERREUR DE COUVERTURE LORS DU RECENSEMENT DE LA POPULATION DE 1996

Claude Julien¹

RÉSUMÉ

Une des principales utilisations du recensement est de servir de base à l'estimation de la taille de chaque province pour l'allocation des transferts fiscaux. Depuis 1991, les estimations des études de couverture sont employées pour corriger les chiffres de base du recensement. Il devient ainsi encore plus important d'améliorer ces études pour réduire les sources d'erreur échantillonnale et non-échantillonnale. L'Étude de la contre-vérification des dossiers est l'étude de couverture la plus importante en termes de volume et d'impact sur les transferts fiscaux. Cet article présente comment nous avons amélioré cette étude pour produire avec les mêmes ressources des estimations plus précises tout en réduisant les sources de biais potentiel.

MOTS CLÉS: erreur de couverture; recensement; erreur non-échantillonnale; appariement.

1. INTRODUCTION

Depuis 1961, Statistique Canada publie des estimations de l'erreur de couverture pour les recensements de la population tenus à tous les cinq ans. Ces estimations sont produites en employant la technique de contre-vérification des dossiers (CVD). Avant 1991, ces estimations informaient les gestionnaires du recensement et les principaux utilisateurs de données de la qualité de la couverture. En 1991, Statistique Canada a décidé d'inclure ces estimations dans le programme d'estimation de la population. Pour ce faire, Statistique Canada devait réduire les sources d'erreurs affectant la qualité des CVD de 1981 et 1986 décrites dans Burgess (1988). En 1991, Statistique Canada a presque doublé la taille de l'échantillon de la CVD, introduit des améliorations importantes dans le traitement des données et produit des estimés officiels du surdénombrement pour la première fois.

Les estimations de l'erreur de couverture ont maintenant un impact direct sur le partage des transferts fiscaux du gouvernement fédéral aux provinces. Cette nouvelle utilisation des estimations a eu deux conséquences. Tout d'abord, les estimations de 1991 ont été analysées de beaucoup plus près et toutes les sources d'erreur échantillonnale et non-échantillonnale ont été identifiées comme importantes. Ensuite, vu la sensibilité des formules de transfert, les sources d'erreur doivent être réduites davantage. Ainsi, pour la CVD de 1996, le plan d'enquête a été modifié pour produire des estimations plus précises au niveau des provinces avec la même taille d'échantillon qu'en 1991. En contrepartie, la composition démographique de l'échantillon de 1996 risque d'augmenter la non-réponse, et par le fait même, l'erreur non-échantillonnale. Pour réduire ce risque, nous avons amélioré davantage le traitement des données.

Les quatre sections suivantes présentent les sources avérées ou potentielles d'erreur qui caractérisent chaque opération de la CVD. Les quatre principales opérations de la CVD sont: (1) le plan d'enquête, (2) le dépistage des personnes choisies et la collecte de données, (3) la recherche de documents du recensement et la classification des personnes choisies et (4) la pondération et l'analyse. La première opération consiste à réunir plusieurs sources de données pour créer la population qui aurait dû être dénombrée au recensement et à choisir un échantillon parmi ces personnes. La seconde opération consiste à contacter les personnes choisies (PC) après le recensement et leur demander toutes les adresses où elles

¹ Claude Julien, Statistique Canada, 15-C immeuble R.H. Coats, Parc Tunney, Ottawa, Ontario, Canada, K1A 0T6.

auraient pu être dénombrées. La troisième opération consiste à trouver le document du recensement complété à chaque adresse pour déterminer le nombre de fois qu'une PC est effectivement dénombrée. Cette opération peut fournir de nouvelles pistes pour contacter une PC ou peut nécessiter la collecte d'information supplémentaire. La dernière opération consiste à pondérer l'échantillon, à compenser pour la non-réponse, à produire des estimations et à les comparer à des sources indépendantes. Cette opération nous permet d'identifier des problèmes qui pourraient engendrer la révision de la classification d'un groupe de PC.

2. PLAN D'ENQUÊTE

2.1 Base d'échantillonnage

La base d'échantillonnage est constituée à partir des six listes ou bases suivantes:

<i>Recensement</i>	Toutes les personnes dénombrées dans une province au recensement précédent; pour la CVD96, toutes les personnes recensées dans une province en 1991.
<i>Naissances</i>	Toutes les naissances survenues durant la période intercensitaire (base constituée à partir des dossiers de la statistique de l'état civil).
<i>Immigrants</i>	Tous les immigrants reçus pendant la période intercensitaire (base constituée à partir des dossiers de Citoyenneté et Immigration Canada).
<i>Détenteurs de permis</i>	Toutes personnes demeurant au Canada en vertu d'un permis temporaire d'étude, de travail ou d'un ministre, ainsi que les personnes réclamant le statut de réfugié (base constituée à partir des dossiers de Citoyenneté et Immigration Canada).
<i>Personnes non recensées</i>	Échantillon des personnes non dénombrées au recensement précédent ; pour la CVD96, le groupe de personnes qui, d'après la CVD91, n'ont pas été recensées (il n'existe pas de liste complète de ce groupe, mais seulement un échantillon de ces personnes accompagnées de leur poids de sondage final de 1991).
<i>Fichier de santé</i>	Toutes personnes inscrites sur les fichiers de l'assurance-santé des Territoires du Yukon et des Territoires du Nord-Ouest (base employée uniquement pour ces territoires).

Ces bases sont censées comprendre toutes les personnes qui doivent être dénombrées au recensement de 1996 sans doublons ni chevauchement. Royce (1993) rapporte que ces bases couvrent au-delà de 99% la population visée. Pour ainsi dire, la couverture des bases ne constitue pas une source importante d'erreur. Malgré cela, certaines lacunes méritent d'être signalées.

Certaines personnes ne figurent pas sur ces listes. C'est le cas des Canadiens qui étaient à l'étranger au moment du recensement précédent et qui sont revenus au pays depuis. Ce groupe représente environ 0,5% de la population canadienne et n'est inscrit sur aucun registre. Il en est ainsi depuis la première CVD effectuée en 1961. Pour les mêmes raisons, les bases employées pour la CVD96 ne couvrent pas les personnes qui demeuraient dans un territoire lors du recensement précédent et qui ont déménagé dans une province depuis. La taille de ce groupe ce chiffre à moins de 4 500 personnes et leur contribution à l'estimation du sous-dénombrement est inférieure à 0,1%. Bureau, Julien et Provost (1995) justifient la décision d'exclure cette population par la nécessité d'employer une source plus récente pour mesurer l'erreur de couverture dans les territoires et le coût de traiter un échantillon de la base du recensement provenant des territoires.

La plupart des doublons proviennent de la base du recensement. Pour la première fois en 1991, les études de qualité ont permis d'estimer le nombre de doublons présents dans le recensement. Ainsi, nous pourrions utiliser ces estimations pour corriger le poids des unités choisies de la base du recensement et enrayer la contribution des doublons.

Deux sources de chevauchement entre les bases méritent une attention particulière. Tout d'abord, certains immigrants reçus et détenteurs de permis demeuraient au Canada au moment du dernier recensement en vertu d'un permis temporaire. Ces personnes font partie des bases du recensement et des personnes non dénombrées. Cette source d'erreur potentielle a été introduite dans la CVD par l'inclusion, pour la première fois en 1991, de la population de détenteurs de permis dans la population visée par le recensement. Les données administratives de Citoyenneté et Immigration Canada (CIC) nous permettent d'identifier la plupart de ces personnes choisies dans l'échantillon. L'autre source de chevauchement provient des erreurs de classification des CVD précédentes. Une personne classée comme non dénombrée par la CVD précédente alors qu'elle était effectivement dénombrée constitue un chevauchement entre la base des personnes non dénombrées et celle du recensement. Cette source d'erreur est inhérente à la méthodologie de la CVD. Pour la contrer, nous introduisons à chaque CVD des améliorations opérationnelles. Malgré cela, elle sera toujours présente et nous supposons que ses effets sont négligeables et même annulés en grande partie par d'autres sources d'erreurs.

Enfin, quoique la non-réponse complète au recensement constitue un phénomène assez marginal, elle présente pour la CVD une source d'erreur qui risque de prendre de l'ampleur. Dans le cas d'une non-réponse complète au recensement, nous employons des procédures spéciales à la collecte et au traitement des données pour ajouter des personnes par imputation. Le recensement de 1991 a vu le nombre de personnes ajoutées augmenter de 92% par rapport au recensement précédent. Un biais dans le nombre de personnes imputées ou leurs caractéristiques affectent directement la qualité de la base du recensement.

2.2 Sélection de l'échantillon

Le plan de sondage constitue la principale amélioration méthodologique mise en place pour la CVD96. À l'exception de la base des personnes non dénombrées dont l'échantillon nous vient de la CVD91, nous tirons un échantillon de personnes à partir d'un plan d'enquête à un seul niveau avec une stratification démographique et une répartition optimale en fonction du taux historique de sous-dénombrement et la taille des strates. Les CVD précédentes employaient un plan à deux niveaux avec une stratification géographique dans la base du recensement. Il en résulte que le plan de 1996 produira des estimations plus précises, car, comme le montrent Boudreau et Germain (1990) et Royce, Germain, Julien, Dick, Switzer et Allard (1994), l'erreur de couverture est un phénomène davantage démographique que géographique. Grâce à ce plan d'enquête, nous avons sélectionné plus de personnes qui sont difficiles à dénombrer. En observant plus de cas de personnes non dénombrées dans l'échantillon, ce que nous voulons mesurer, nous produirons des estimations plus précises du sous-dénombrement au niveau provincial.

Tableau 1 - Comparaison des tailles d'échantillon

STRATE	proportion de l'échantillon		ratio 96/91	taux de réponse
	CVD96	CVD91		
HOMMES 30-39 NON MAR	3.1%	1.4%	220%	91%
HOMMES 40-59 NON MAR	2.9%	1.3%	220%	93%
HOMMES 20-29 NON MAR	7.8%	4.2%	185%	93%
FEMMES 30-39 NON MAR	2.4%	1.3%	184%	93%
DÉTENTEURS DE PERMIS IMMIGRANTS	2.3%	1.3%	180%	79%
FEMMES 20-29 NON MAR	5.2%	3.0%	173%	95%
HOMMES 40-59 MAR	5.0%	6.8%	74%	98%
FEMMES 15-19	4.4%	6.0%	72%	96%
FEMMES 30-39 MAR	3.4%	5.0%	68%	98%
HOMMES 60+ MAR	2.7%	4.1%	66%	99%
HOMMES 20-29 MAR	1.8%	2.8%	65%	96%
FEMMES 60+ MAR	2.0%	3.2%	63%	99%
FEMMES 40-59 MAR	3.9%	6.4%	62%	98%
FEMMES 20-29 MAR	2.0%	3.7%	53%	97%

Le tableau 1 montre les groupes de personnes dont la taille d'échantillon en 1996 a le plus augmenté ou diminué par rapport à la CVD91. Ce même tableau montre aussi le taux de non-réponse observé à la CVD en 1991 pour ces groupes de personnes. Les personnes qui sont difficiles à dénombrier sont jeunes et mobiles et, en conséquence, difficile à dépister et à classer avec exactitude. Nous sommes en présence du dilemme où nous réduisons l'erreur échantillonnale au risque d'augmenter l'erreur non-échantillonnale. Pour cette raison, nous introduisons quelques améliorations opérationnelles d'envergure pour contrôler la non-réponse et le potentiel d'erreur qui s'y rattache.

3. DÉPISTAGE

Le dépistage est, en terme d'erreur potentielle et ses effets, l'opération la plus importante et la plus difficile à améliorer. À partir des données provenant des sources de sélection, nous devons contacter chaque PC le plus tôt possible après le recensement. Or, certaines sources de données sont vieilles de 5 ans, comme la base du recensement. Le défi est assez considérable. En employant des données administratives et en exploitant le talent des interviewers dans nos bureaux régionaux, nous réussissons, d'une CVD à l'autre, à contacter plus de 96% des PC.

3.1 La problématique des personnes non dépistées

Malgré cet excellent résultat, la faible proportion de personnes non dépistées peut provoquer des erreurs significatives dans l'estimation d'un taux de sous-dénombrement de 3%. Le problème est d'établir le taux de sous-dénombrement qu'on doit « imputer » aux personnes non dépistées? On s'attend à ce que leur taux soit plus élevé que chez les personnes dépistées. En d'autres mots, une personne non dénombrée est aussi une personne difficile à dépister. La CVD91 a « imputé » un taux de 12,7% aux personnes non dépistées contre un taux estimé de 9,3%² chez les personnes dépistées. Cet écart est-il juste? Comme le rapporte Burgess « S'il n'est pas exact, il pourrait y avoir distorsion dans les estimations provinciales du sous-dénombrement et un biais dans les estimations globales du sous-dénombrement ». Burgess soumet l'hypothèse que les personnes non dépistées ont tendance à avoir déménagé de province depuis leur sélection (migrant inter-provincial). Or, les migrants inter-provinciaux présentent un taux de sous-dénombrement (9,9 %) supérieur à l'ensemble de la population (3,4 %). Si l'hypothèse de Burgess est vraie, la CVD aurait tendance à sous-estimer le sous-dénombrement.

Le tableau 2 montre le résultat d'une étude où nous avons simulé l'impact de divers taux de sous-dénombrement imputés aux personnes non dépistées. La première partie du tableau présente l'impact sur l'estimation du nombre de personnes non dénombrées. Un taux imputé entre 15% et 20% engendre un biais supérieur à l'écart-type seulement pour l'estimation de l'ensemble des 10 provinces. Il faut un taux imputé supérieur à 21,4% (au lieu du 12,7% effectivement imputé) pour que l'estimation dans les plus grandes provinces (QC, ON, AB et CB) diffère de plus d'un écart-type de l'estimation finale de la CVD91. La seconde partie du tableau présente l'impact sur l'estimation de la part relative de la province. Celle-ci correspond au compte du recensement ajusté par l'estimation de personnes non dénombrées dans la province divisé par la somme de ceci pour l'ensemble des 10 provinces. On voit que même avec un taux imputé de plus de 33%, l'impact sur la part relative est inférieur à l'écart-type. Ces résultats confirment ce que Burgess a rapporté à savoir qu'un taux imputé légèrement supérieur peut occasionner un biais important au niveau national. Par contre, au niveau provincial, le biais est faible par rapport à l'erreur échantillonnale et, par conséquent, a peu d'impacts sur les transferts fiscaux qui sont tributaires de la part relative de chaque province.

² Lors de la CVD91, près de la moitié des PC dénombrées ont été classées sans dépistage lors d'une opération spéciale intégrée au traitement du recensement. Ceci explique les taux élevés en comparaison au taux global de 3,4%. Cette opération n'aura pas lieu lors de la CVD96.

Tableau 2 - Impact du taux imputé aux PC non dépistées

Impact sur l'estimation du nombre de personnes non recensées								
Province	CVD91		taux imputé aux PC non dépistées					
	estimation	écart-type	13.7%	14.7%	15.6%	21.4%	28.0%	33.2%
Terre-Neuve	17747	1771	33	64	95	293	530	730
Île-du-Prince-Édouard	2928	306	57	61	65	89	117	140
Nouvelle-Écosse	26072	3379	11	73	132	500	913	1240
Nouveau-Brunswick	32360	3260	123	243	360	1102	1986	2725
Québec	281444	14487	1443	2840	4196	12681	22507	30471
Ontario	521880	31328	3423	6727	9921	29661	51963	69595
Manitoba	32119	4125	149	293	433	1309	2321	3139
Saskatchewan	32591	3329	129	254	374	1121	1971	2653
Alberta	87826	7202	705	1393	2063	6341	11469	15759
Colombie-Britannique	144016	8441	915	1803	2667	8109	14479	19686
Total	1178983	36536	6988	13751	20306	61206	108256	146138

Impact sur la part relative de chaque province (en %)								
Province	CVD91		taux imputé aux PC non dépistées					
	estimation	écart-type	13.7%	14.7%	15.6%	21.4%	28.0%	33.2%
Terre-Neuve	2.1	0.007	0.000	-0.001	-0.001	-0.003	-0.006	-0.008
Île-du-Prince-Édouard	0.5	0.001	0.000	0.000	0.000	-0.001	-0.001	-0.002
Nouvelle-Écosse	3.3	0.012	-0.001	-0.001	-0.002	-0.005	-0.009	-0.012
Nouveau-Brunswick	2.7	0.012	0.000	0.000	-0.001	-0.002	-0.003	-0.004
Québec	25.3	0.047	-0.001	-0.002	-0.003	-0.010	-0.017	-0.023
Ontario	37.4	0.091	0.003	0.006	0.008	0.024	0.041	0.053
Manitoba	4.0	0.015	0.000	-0.001	-0.001	-0.004	-0.007	-0.009
Saskatchewan	3.6	0.012	0.000	-0.001	-0.001	-0.004	-0.007	-0.009
Alberta	9.3	0.026	0.000	0.000	0.001	0.002	0.005	0.008
Colombie-Britannique	12.1	0.030	0.000	0.001	0.001	0.003	0.005	0.007
Total								

Note: le taux imputé à la CVD91 est 12.7%

3.2 Le contrôle du taux de non-dépistage

L'effort pour réduire le nombre de personnes non dépistées commence dès la sélection de l'échantillon. À cette opération, nous consignons l'information pertinente au dépistage de la PC et de toute personne demeurant avec la PC ou associée à la PC (par exemple, le parrain d'un immigrant reçu). Cette information est très utile car il est plus facile de dépister un ménage qu'une personne seule. Les bases du recensement, des naissances et des personnes non dénombrées nous fournissent naturellement des ménages. Par contre, extraire des ménages des données administratives de CIC et des fichiers des territoires demande plus d'efforts.

L'étape cruciale du dépistage consiste à appairer les PC et les membres de leur ménage à plusieurs sources de données administratives pour trouver leur adresse la plus à jour avant le recensement. Les sources de données consultées sont les données fiscales, les données de CIC et les annuaires téléphoniques disponibles sous format électronique. Le tableau 3 montre la proportion de PC pour lesquelles au moins un membre du ménage a été apparié aux données fiscales. La CVD96 présente de meilleurs résultats pour les PC des bases du recensement, des naissances et des personnes non dénombrées. Pour la première fois, nous avons aussi apparié les immigrants. De plus, les données administratives de CIC nous donnent en 1996 des adresses très récentes pour la plupart des détenteurs de permis.

Après avoir obtenu les adresses les plus récentes possible, nous envoyons l'information aux bureaux régionaux où des interviewers tentent de contacter les PC par téléphone à une des adresses fournies ou en les dépistant avec des sources publiques acquises par chaque bureau. Les améliorations que nous avons apportées au plan de sondage et aux opérations de sélection nous permettent de débiter cette opération plus tôt que lors des CVD précédentes. C'est surtout le cas pour les échantillons des immigrants, des détenteurs de permis et des territoires. Ceci aura, nous l'espérons, un impact positif sur le taux de

dépistage. De plus, en parallèle au dépistage sur le terrain, nous continuons à consulter les données administratives plus récentes (par exemple, les données fiscales de 1995 et 1996) pour obtenir d'autres adresses et les faire parvenir, au besoin, aux bureaux régionaux.

3.3 La collecte

Les interviewers tentent de contacter directement chaque PC ou un parent ou un conjoint pour compléter un questionnaire d'enquête. Lors de l'entrevue le répondant rapporte toutes les adresses où la PC aurait pu être dénombrée et les caractéristiques du ménage dans lequel la PC vivait. Ces adresses comprennent la résidence actuelle, l'adresse le jour du recensement, l'adresse précédente (si la PC a déménagé récemment), des adresses où la PC a demeuré temporairement durant la période de collecte du recensement et l'adresse d'un parent ou un ami de la PC qui aurait pu l'inclure sur son questionnaire du recensement. En combinant la source de sélection, les sources administratives et la collecte, nous pouvons avoir jusqu'à 10 adresses par PC. En moyenne, nous prévoyons en avoir au moins trois. Les caractéristiques du ménage comprennent le nom, le sexe et la date de naissance des personnes qui vivaient avec la PC à son adresse habituelle le jour du recensement. Finalement, nous obtenons des informations comme l'état matrimonial, la langue parlée et le type de logement pour établir un meilleur profil des personnes non dénombrées. La prochaine opération consiste à saisir cette information et initier la recherche du document du recensement complété à chaque adresse obtenue.

Tableau 3 - Appariement aux données fiscales

	Taux d'appariement	
	CVD91	CVD96
BASES		
RECENSEMENT	85%	92%
NAISSANCES	73%	80%
NON RECENSÉES	70%	80%
IMMIGRANTS	Pas tenté	72%
DÉTENTEURS DE PERMIS	Pas tenté	Pas tenté
STRATES DE LA BASE DU RECENSEMENT		
HOMMES 30-39 NON MAR	76%	87%
HOMMES 40-59 NON MAR	72%	86%
HOMMES 20-29 NON MAR	85%	92%
FEMMES 30-39 NON MAR	75%	89%
FEMMES 20-29 NON MAR	87%	93%
HOMMES 40-59 MAR	86%	94%
FEMMES 15-19	89%	94%
FEMMES 30-39 MAR	91%	95%
HOMMES 60+ MAR	82%	94%
HOMMES 20-29 MAR	89%	94%
FEMMES 60+ MAR	78%	93%

4. RECHERCHE ET CLASSIFICATION

L'opération de la recherche et classification consiste à traiter l'information obtenue au dépistage et à la collecte pour déterminer si la PC devait être dénombrée et, le cas échéant, le nombre de fois. Parmi les personnes ne devant pas être dénombrées figurent les personnes décédées et les personnes ayant quitté le Canada avant le jour du recensement. Pour les autres PC, il faut consulter le document du recensement correspondant à chaque adresse disponible. Par document du recensement, nous désignons les registres de visites, les questionnaires dûment ou partiellement complétés et les formulaires indiquant la raison d'une non-réponse complète (non-contact, refus, etc.).

4.1 La problématique de la recherche et la classification

Pour classer une PC comme dénombrée, il faut la trouver « clairement » recensée sur un des documents consultés. Pour classer une PC comme non dénombrée, il faut être certain d'avoir trouvé les documents correspondant à toutes les adresses où la PC aurait pu être dénombrée sans qu'elle y soit « clairement »

dénombrée. Une PC est «clairement» dénombrée à une adresse lorsque son nom ou ses caractéristiques (sexe et date de naissance) figurent sur un questionnaire.

Par opposition, une PC ne peut pas être «clairement» dénombrée à une adresse où une procédure du recensement a imputé un nombre de personnes à cause d'une non-réponse complète. Dans ce cas, ainsi que d'autres cas où l'information et les résultats obtenus ne sont pas assez précis pour en arriver à un classement définitif, la PC est «non classée» et contribue à augmenter la non-réponse de la CVD, tout comme les personnes non dépistées. La proportion des PC non classées est faible (1,1%), mais ces personnes posent le même problème que les PC non dépistées, à savoir quel taux de sous-dénombrement doit-on leur «imputer»? La CVD91 leur a «imputé» un taux de 12.0% contre un taux estimé de 9.5% pour les personnes classées.

Les erreurs de classification engendrent un biais. D'une part, il est difficile de déclarer à tort une PC comme dénombrée. D'autre part, nous ne pouvons jamais avoir la certitude qu'une PC est non dénombrée à moins de consulter tous les documents du recensement. Il se peut que le répondant ne nous donne pas toutes les adresses ou que nous n'ayons pas traité toutes les adresses correctement. Par conséquent, la classification est caractérisée par un biais incontournable que nous devons réduire le plus possible.

La réduction de la non-réponse et de l'erreur de classification posent un défi opérationnel considérable. En effet, nous prévoyons avoir à rechercher le document correspondant à plus de 150 000 adresses obtenues au dépistage et à la collecte. Cette opération est rendue particulièrement difficile par le fait que les noms et les adresses ne sont pas saisis au recensement. Nous relevons ce défi en employant un traitement informatique qui réduira le nombre de documents à consulter pour classer une PC. Ceci nous permettra de consacrer plus de temps et de ressources à traiter correctement les PC éventuellement classées comme non dénombrées ou non classées.

4.2 Le codage géographique et l'appariement

Le traitement informatique comporte deux étapes, le codage géographique et l'appariement. Le codage géographique consiste à assigner à chaque adresse une zone de recherche où elle devrait être située. Cette zone est composée de 1 à 10 secteurs de dénombrement et provient essentiellement de la conversion du code postal en un ensemble de secteurs. Cette étape a été mise en place pour la CVD91 et améliorée pour la CVD96.

L'appariement consiste à appairer le ménage dans lequel vivait la PC à tous les ménages présents dans la base de données du recensement dans la zone de recherche identifiée par le codage géographique. En employant seulement les caractéristiques (sexe et date de naissance) de la PC et des membres du ménage, l'appariement trouve tous les ménages de la base qui ressemblent à celui de la PC. Lorsqu'un des ménages de la base ressemble «fortement» à celui de la PC et que la PC «participe» à cette ressemblance, nous pouvons classer la PC comme dénombrée sans consulter le questionnaire du recensement du ménage en question pour confirmer la présence de la PC.

Dans une étude, dont le but premier était d'évaluer une façon de déceler le surdénombrement dans le recensement, Bernier (1995) a montré que deux ménages de la base présentant plusieurs personnes avec le même sexe et date de naissance étaient nécessairement composés des mêmes personnes. Par exemple, deux ménages ayant au moins trois personnes en commun sont nécessairement le même ménage dénombré deux fois. Une autre façon d'interpréter ce résultat est de déclarer que les ménages d'au moins deux personnes sont uniques, en terme de combinaison sexe et date de naissance des membres, dans une zone aussi restreinte que celle produite par le codage géographique. Ainsi, dans la CVD96, nous classerons automatiquement un PC comme dénombrée lorsque nous appairerons ses caractéristiques ainsi que celles des membres de son ménage au même ménage de la base de données du recensement.

Nous avons appliqué cette méthode aux données de la CVD91. Cette méthode permet de classer plus de la moitié des PC dénombrées automatiquement sans consulter de documents du recensement. Par contre, la méthode n'est pas sans erreur. En effet, la méthode déclare deux personnes effectivement non dénombrées comme dénombrées. Puisqu'il y avait 2 341 personnes non dénombrées, le taux d'erreur est inférieur à

0,1% et négligeable en comparaison aux bénéfices introduits par la méthode qui sont la classification consistante, systématique et ordonnée ainsi qu'une meilleure utilisation des ressources.

En plus, l'appariement identifie des ménages de la base quasi semblables à celui de la PC, mais pas assez pour une classification automatique. Ceci permet de réduire la tâche de vérification à celle de consulter le questionnaire du recensement de un ou deux ménages. La stratégie qui consiste à appairer le ménage de la PC, et non pas seulement la PC, à la base du recensement constitue l'amélioration la plus importante mise en place dans le traitement des données de la CVD96. Elle permettra de traiter très rapidement plusieurs PC avec exactitude et nous laissera plus de temps et de ressources à effectuer, vérifier et corriger le traitement des PC éventuellement non dénombrées ou non classées.

4.3 L'appariement monstre et recherches supplémentaires

Cette opération, nouvelle à la CVD96, supplémente le codage géographique et l'appariement. Elle consiste à appairer le ménage de la PC à la base de données du recensement pour y identifier tous les ménages ayant au moins deux personnes en commun avec celui-ci. Cette opération emploie un algorithme décrit dans Julien et Mayda (1995). Elle permet de classer des PC comme dénombrées à des adresses qui ne sont pas données à la collecte ou à des adresses traitées incorrectement à la recherche. L'appariement monstre est aussi une méthode simple et peu coûteuse pour résoudre des adresses et trouver de nouvelles pistes pour classer la PC correctement. Elle permet ainsi de réduire la non-réponse et l'erreur de classification tout en évitant de recourir à un nouveau contact auprès de la PC.

Avant la classification finale, les personnes non dénombrées après les recherches déjà décrites font l'objet de recherches supplémentaires. Nous tentons d'obtenir d'autres adresses des sources administratives. Les résultats de l'appariement monstre peuvent nous fournir de nouvelles pistes en vue de contacter les PC. Finalement, toutes les PC non dénombrées et certaines PC non classées font l'objet d'un second contact où nous confirmons et précisons leur adresse le jour du recensement et tentons d'obtenir d'autres adresses. Toute cette nouvelle information est traitée avec la plupart des opérations déjà décrites.

En résumé, l'opération de recherche et classification consiste à traiter un volume considérable d'information dans des conditions difficiles. Il est important d'effectuer cette opération le plus correctement et complètement possible. D'une part, la classification de personnes comme non dénombrées alors qu'elles sont effectivement dénombrées mène à surestimer le sous-dénombrement. D'autre part, l'impossibilité de classer définitivement certaines personnes peut mener à sous-estimer le sous-dénombrement. Pour réduire l'occurrence de ces erreurs, nous avons développé des procédures automatiques de classification qui effectuent rapidement le traitement de plusieurs PC avec un risque d'erreur négligeable. Avec les mêmes ressources qu'en 1991, ceci nous permet de consacrer plus d'effort à classer correctement les autres PC. Par contre, la non-réponse complète au recensement rend quelques PC impossibles à classer et augmente la non-réponse de la CVD sans qu'on puisse y faire quoique que ce soit. Si la non-réponse complète au recensement augmente davantage, il faudra envisager de nouvelles méthodes pour traiter cette situation.

5. PONDÉRATION ET ANALYSE

La pondération comporte trois étapes. Lors de la sélection, nous attribuons à chaque PC un poids initial équivalent à l'inverse de sa fraction de sondage. Pour la base des personnes non dénombrées, le poids initial correspond au poids final de la CVD précédente. À cette étape, nous sélectionnons l'échantillon en cinq répliques pour permettre l'estimation de la variance. Par la suite, ce poids subit deux corrections: l'une pour tenir compte de la non-réponse, l'autre pour améliorer la représentativité de l'échantillon par rapport aux bases dont il est tiré (poststratification).

Dans les CVD précédentes, ces corrections étaient effectuées à l'intérieur de groupes définis en fonction du plan de sondage (bases et strates) et, si possible, des caractéristiques démographiques et de mobilité tout en respectant la division en répliques. Le fait que la base du recensement était stratifiée

géographiquement limitait grandement les regroupements démographiques qu'on pouvait y effectuer. La stratification entièrement démographique du plan de sondage de la CVD96 permet des regroupements plus fins pour corriger la non-réponse et abolit pratiquement toute nécessité pour la seconde correction. Nous investiguons actuellement des façons d'employer davantage les résultats du dépistage et de la recherche pour effectuer de meilleures corrections.

Les poids finaux sont employés pour estimer le sous-dénombrement, bien sûr, mais aussi le nombre de personnes dénombrées, décédées, ayant quitté le Canada, etc. Ces estimations sont comparées à des sources externes pour identifier les problèmes potentiels. Les estimations de sous-dénombrement sont comparées aux erreurs de fermeture dérivées des projections démographiques et aux estimations des CVD précédentes. Le profil des personnes non dénombrées est analysé conjointement avec la division de la démographie. En 1991, nous avons procédé à la révision complète des personnes non dénombrées au Nouveau-Brunswick et pour certains groupes d'âge. Dans le cas du Nouveau-Brunswick, les estimations de la CVD se sont avérées exactes, alors que l'estimation pour les filles de 0 à 4 ans a été ajustée pour assurer une certaine cohérence démographique (rapport homme/femme).

Les estimations de personnes dénombrées par province et par groupe démographique sont comparées aux comptes du recensement. Ceci permet d'analyser la pondération et d'évaluer si les corrections pour la non-réponse distribuent assez bien le poids des personnes non dépistées et les personnes non classées vers les personnes dénombrées et les personnes non dénombrées. Historiquement, la CVD a bien estimé la population dénombrée par province. Au niveau des groupes démographiques, la comparaison devient difficile à cause de l'erreur échantillonnale plus grande et l'augmentation de la non-réponse partielle et totale au recensement qui crée des distorsions dans la distribution démographique de la population dénombrée.

Les estimations de personnes décédées par province sont comparées aux comptes rapportés par les statistiques de l'État civil. Des estimations suspectes mènent à la révision de l'information et à une réévaluation des résultats obtenus. À cette étape, nous procédons à l'appariement de la personne décédée au registre des décès pour confirmer le résultat de la classification. De même, les estimations du nombre de personnes ayant quitté le pays sont comparées aux estimations basées sur des modèles démographiques.

En résumé, la pondération est une opération importante, car elle sert à corriger les poids pour compenser pour la non-réponse. Il est essentiel que l'ajustement s'effectue à l'intérieur de groupes formés de personnes présentant des caractéristiques similaires telles la base de sondage, l'âge, l'état matrimonial et le sexe qui influent sur leur dénombrement. Toutes ces caractéristiques font partie du plan de sondage de la CVD96 qui permettra ainsi un regroupement plus approprié qu'en 1991. Les procédures employées au dépistage, à la recherche et à la classification permettent de déceler et de traiter des problèmes au niveau des micro-données, alors que la comparaison des estimations de la CVD avec des sources externes permet de déceler des problèmes au niveau des macro-données. Cette analyse peut mener à la révision complète de toute la documentation accumulée pour un groupe de PC pour corriger ou confirmer les estimations de la CVD.

6. CONCLUSION

Depuis 1961, les études de couverture du recensement de la population informent les gestionnaires du recensement et les utilisateurs sur la qualité de la couverture. Depuis 1991, ces études ont aussi un impact direct sur les finances des provinces puisqu'elles produisent des estimations qui sont utilisées pour corriger les comptes du recensement et produire des estimations de la population de chaque province. En conséquence, il est essentiel de réviser la méthodologie et les procédures de traitement pour réduire davantage les erreurs échantillonnales et non dues à l'échantillonnage. Pour la CVD de 1996, nous avons poursuivi les améliorations entamées dans la CVD de 1991. D'une part, nous avons modifié le plan de sondage pour obtenir des estimations plus précises avec la même taille d'échantillon. D'autre part, nous avons amélioré les procédures de dépistage, de recherche et de classification pour réduire la non-réponse et les autres erreurs de traitement pouvant occasionner des biais. Nous dirigeons présentement nos recherches vers la pondération et l'ajustement pour la non-réponse. Nous chercherons à exploiter le

nouveau plan de sondage et les résultats du traitement pour former de meilleurs regroupements visant à distribuer le poids des PC non dépistées et non classées.

7. BIBLIOGRAPHIE

Bernier, J. (1995) « Étude pilote: estimation du surdénombrement détecté par appariement automatique », rapport interne.

Boudreau, J.-R. et Germain M.-F. (1990) *Guide à l'intention des utilisateurs des données sur la qualité du recensement de 1986: Couverture*, Statistique Canada, Catalogue 99-135F.

Bureau, M., Julien, C. et Provost, M. (1995) « Les études de mesure de l'erreur de couverture dans les territoires en 1996 », rapport interne.

Burgess, R. D. (1988) « Évaluation des estimations du sous-dénombrement obtenus par la contre-vérification des dossiers du recensement du Canada », *Techniques d'enquêtes*, Vol. 14, No. 2, Statistique Canada, pp. 147-167.

Julien, C. et Mayda, M. (1995) « Improving Census Coverage Error Measurement Through Automated Matching », *Proceedings of the Survey Methods Section, American Statistical Association*, pp. 849-854.

Royce, D. (1993) « Comments on Documents from Bureau de la Statistique du Québec Concerning Adjustment of the Population Estimates for Net Undercoverage », rapport interne.

Royce, D., Germain, M.-F., Julien, C., Dick, P., Switzer, K. et Allard, B. (1994) *Couverture: Rapport technique du recensement de 1991*, Statistique Canada, Catalogue 92-341F.

QUEL EST LE RÔLE DE L'ANALYSE DÉMOGRAPHIQUE DANS LE RECENSEMENT DE 2000 DES ÉTATS-UNIS?

J. Gregory Robinson¹

RÉSUMÉ

L'analyse démographique (AD) est un programme américain de mesure et d'évaluation de la couverture bien structuré. L'AD a servi d'étalon pour jauger les tendances relatives à la couverture lors des derniers recensements et les variations de couverture selon l'âge, le sexe et la race à l'échelle nationale. Nous examinerons ici le rôle de l'analyse démographique dans le Recensement de 2000. Deux questions se posent :

* L'AD ne devrait-elle servir qu'à évaluer la couverture en vue de confirmer les estimations obtenues par sondage (CensusPlus ou système double) et dont on se sert dans le cadre du Programme de mesure intégrée de la couverture (PMIC)?

Ou

*Devrait-on officiellement incorporer l'AD aux estimations par sondage dans le cadre du Programme de mesure intégrée de la couverture (PMIC), en tirant parti des avantages particuliers de cette approche ?

Le rôle de l'AD devrait illustrer un équilibre entre les forces et les faiblesses de la méthode démographique et des estimations de couverture effectuées par sondage. Nous croyons que l'analyse démographique peut jouer un rôle important, plus vaste, au niveau de l'évaluation, lors du Recensement de 2000. Elle peut aussi améliorer le PMIC là où elle donne de bons résultats et où les estimations par sondage laissent à désirer, soit 1) pour mesurer le sous-dénombrement chez les hommes adultes de race noire et 2) pour produire des estimations détaillées selon l'âge, le sexe et la race, cohérentes sur les plans longitudinal et interne. En incorporant les résultats de l'AD au PMIC de 2000, on pourra concilier les variations âge-sexe-race entre les estimations AD et les estimations par sondage avant la diffusion des chiffres finaux du recensement.

MOTS CLÉS : Analyse démographique; évaluation de la couverture; sous-dénombrement.

1. INTRODUCTION

Un des objectifs du Recensement de 2000 est de réduire le sous-dénombrement différentiel et le coût du recensement par l'échantillonnage et l'estimation. Dans la mesure du possible, nous recourons aux techniques statistiques et aux dossiers administratifs pour estimer le nombre et le genre de personnes non recensées. Cette activité s'effectuera dans le cadre du Programme de mesure intégrée de la couverture (PMIC). Les personnes non recensées seront ajoutées aux données du recensement afin d'obtenir un «chiffre unique» avant la date de diffusion du 31 décembre 2000.

Le PMIC a été testé pour la première fois lors du recensement expérimental de 1995. Deux techniques de mesure de la couverture avaient alors été mises à l'épreuve - CensusPlus (CP) et l'estimation par système double (ESD). Ces techniques supposent le sondage d'un échantillon où les répondants d'une enquête indépendante sont individuellement appariés à ceux du recensement (lire Mulry et Singh, 1995, pour la description des méthodes CensusPlus et ESD). Lors du programme de mesure de la couverture de 1990, les estimations par sondage reposaient sur l'Enquête postcensitaire (EPC) (lire Hogan, 1993), une technique d'estimation à système double.

Le Census Bureau poursuit un autre programme de mesure et d'évaluation de la couverture, l'analyse démographique. L'analyse démographique (AD) est une approche «macro-analytique» à la quantification de la couverture, en vertu de laquelle on estime le sous-dénombrement net par analyse, en comparant des ensembles de

¹J. Gregory, chef, Population Analysis and Evaluation Staff, Population Division, United States Bureau of the Census, Washington, DC 20233-8800.

données agrégés. Cette approche diffère fondamentalement des estimations par sondage qui correspondent plutôt à une approche «micro-analytique».

L'analyse démographique s'appuie lourdement sur l'agrégation des enregistrements administratifs, essentiellement indépendants du recensement. On estime la population de moins de 65 ans grâce à l'équation que voici :

$$\text{Population} = \text{Naissances} - \text{Décès} + \text{Immigrants} - \text{Émigrants}$$

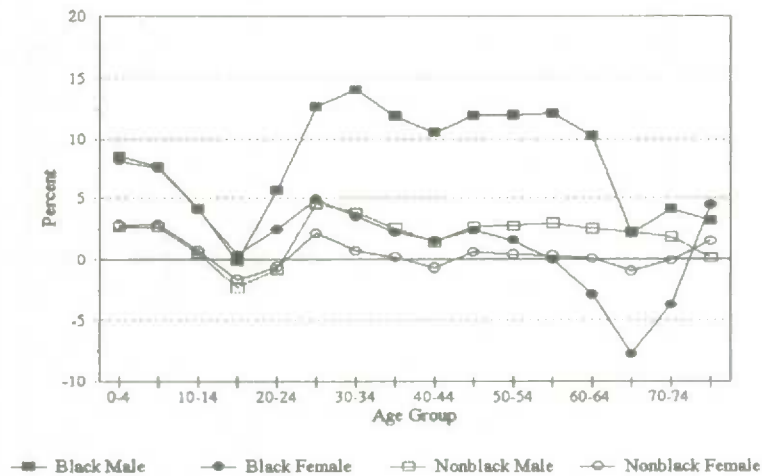
Les données agrégées de l'assurance-maladie servent à estimer la population de 65 ans et plus.

Le processus comporte une série d'hypothèses sur la complétude des données administratives servant à la production des estimations démographiques. De plus, faute de dossiers sur certains membres de la population tels les immigrants illégaux, l'importance de certains groupes doit être estimée (lire Robinson et coll., 1993, pour plus de détails sur les résultats démographiques de 1990, et Himes et Clogg, 1993, pour un excellent aperçu de l'application des méthodes d'analyse démographique).

L'évaluation de la couverture par l'analyse démographique s'est améliorée au fil des ans. Les estimations démographiques nationales constituent désormais des valeurs de référence servant à évaluer les variations de la couverture selon l'âge, le sexe et la race. La figure 1 illustre les taux de sous-dénombrement démographiques en 1990. On constate que le sous-dénombrement était relativement élevé pour les enfants et les hommes de race noire. La tendance la plus notable concerne les hommes de 25 à 64 ans de race noire, pour lesquels le sous-dénombrement se situe entre 10 et 15 %. Un des grands objectifs du Recensement de 2000 est d'atténuer les écarts de ce genre. Dans le contexte du présent symposium, il s'agit d'une erreur non due à l'échantillonnage que l'on essaie de corriger.

Figure 1. Sous-dénombrement net selon la race, le sexe et l'âge en 1990

Figure 1. Percent Net Undercount: 1990 by Race, Sex, and Age



2. RÔLE ÉVENTUEL DE L'AD À L'AN 2000

2.1 Évaluation ou mesure de la couverture

Comment pourrait-on recourir à l'analyse démographique afin de produire un «chiffre de population unique» lors du Recensement de 2000? L'AD peut être utilisée d'une ou deux façons lors du recensement :

(1) Comme instrument pour évaluer la couverture, en complément aux estimations CensusPlus ou ESD du PMIC.

En 1990, on s'est servi de l'AD dans le cadre du programme d'évaluation de la couverture, c'est-à-dire pour déterminer la qualité des résultats de l'EPC et obtenir d'importantes valeurs de référence historiques (1940-1980) qui ont permis d'établir la complétude de la couverture. Les recherches des quatre dernières années ont révélé que l'AD peut intervenir davantage dans l'évaluation du Recensement de 2000, notamment par la production de valeurs de référence infranationales (lire Robinson, 1994 et Robinson et Kobilarcik, 1995).

(2) Comme instrument pour évaluer la couverture et comme programme de mesure «active» de la couverture, en vertu duquel les estimations démographiques de la couverture seraient intégrées aux estimations CensusPlus ou ESD.

Le «meilleur jeu» de valeurs servirait d'étalon au PMIC pour la production du chiffre de population unique du Recensement de 2000. En 1990, on s'était servi exclusivement des estimations de l'EPC pour mesurer la couverture, en prévision d'un rajustement éventuel des chiffres du recensement. On n'avait pas recouru aux estimations de l'AD car à l'époque, on croyait que les inconvénients de cette méthode (aucune ventilation géographique, incertitude des estimations) en dépassaient les avantages (indépendance, cohérence interne).

2.2 Forces et limites de l'AD

L'AD ne devrait-elle servir qu'à l'évaluation en 2000 ou devrait-elle aussi jouer un rôle plus direct par son intégration aux activités de mesure de la couverture du PMIC? La décision dépendra de la manière dont on parvient à minimiser les limites de cette méthode et à en maximiser les points forts. Dans la revue qui suit, nous soulignerons les forces et les limites qui ont changé depuis 1990 en vue d'appuyer l'intégration de l'AD.

2.2.1 Limites de l'AD

1. **Absence de ventilation géographique**--On ne possède pas d'estimations AD indépendantes détaillant l'âge, le sexe et la race sous le niveau national. Pour mesurer la couverture à l'an 2000, les estimations par sondage resteraient le véhicule principal pour les estimations infranationales du PMIC.

Maintes recherches ont été entreprises depuis 1990 en vue d'élaborer une série de valeurs de référence infranationales pour la couverture selon l'AD, notamment à l'égard des personnes de moins de 18 ans et de 65 ans et plus dans les États et les grands comtés. Les données sur les naissances et les décès sont faciles à obtenir pour les plus jeunes, tandis que les registres scolaires permettent de mesurer la population scolaire et d'estimer la migration. Les dossiers de l'assurance-maladie constituent par ailleurs une excellente source indépendante de données pour la population de 65 ans et plus. L'analyse des ratios hommes-femmes nous donne un indice sur les variations de couverture pour les personnes de 18 à 64 ans (lire Robinson, 1994). Enfin, nous avons entrepris l'élaboration d'un programme qui servira à estimer les unités de logement (dans les États et les comtés) et dont les résultats pourront éventuellement être intégrés aux estimations démographiques en vue de les consolider.

Il existe donc une nouvelle dimension géographique au programme démographique qui pourrait devenir un important outil d'évaluation à l'an 2000, en complément des activités d'enquête du PMIC. L'AD nous a permis d'évaluer les résultats CensusPlus et ESD aux sites de recensement expérimental de 1995, à Oakland (Californie), à Paterson (New Jersey) et dans six paroisses rurales de la Louisiane (Robinson, 1996).

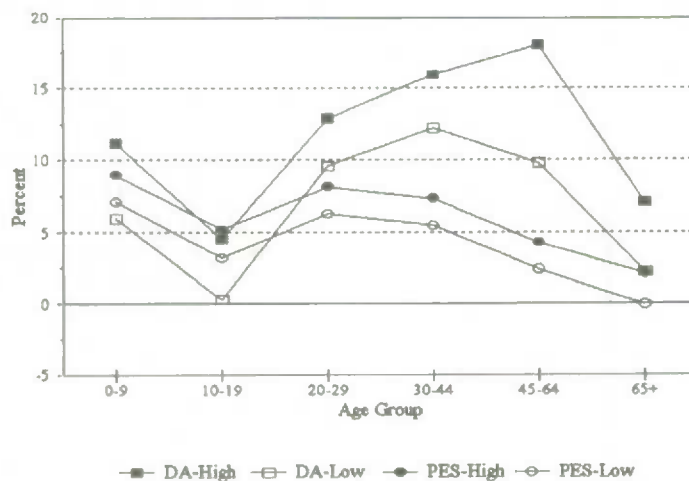
2. **Ventilation limitée pour la race et l'ethnie**--Les principales catégories de l'AD à l'égard de la race sont «race noire» et «autre race». Même si les recherches en vue d'obtenir une estimation de la population d'origine hispanique et asiatique par l'AD se poursuivent, les mesures ne seraient pas aussi fiables que pour les deux groupes précédents. Les estimations CensusPlus/ESD serviraient donc d'étalon pour la mesure de la couverture des personnes d'ascendance hispanique, asiatique ou amérindienne (ainsi que d'autres classifications importantes, selon le mode d'occupation de propriété).

3. **Incohérences dans la classification des races**--Les estimations AD du sous-dénombrement net seront biaisées si les personnes qu'on range parmi les gens de race noire se retrouvent dans un autre groupe ethnique au recensement. Il faut entreprendre plus de recherches pour évaluer le degré d'incohérence et voir comment on pourrait atténuer le plus possible cette «erreur de classification». D'autre part, on doit étudier les conséquences de l'addition d'un groupe multiracial dans la question sur la race du Recensement de 2000.

4. **Incertitude au sujet des estimations AD**--La principale préoccupation qu'ont soulevée les estimations AD en 1990 était l'incertitude des mesures de sous-couverture. Les estimations AD de 1990 étaient accompagnées pour la première fois de mesures statistiques de l'incertitude (Das Gupta, 1991). On a ainsi appris que les estimations AD peuvent souffrir d'une incertitude considérable quant au niveau de sous-dénombrement obtenu (la figure 2 donne les intervalles de confiance à 95 % des estimations AD et EPC du sous-dénombrement de 1990 pour les hommes de race noire). Il est manifeste que les estimations démographiques du sous-dénombrement des hommes de race noire en pourcentage restent relativement élevées, peu importe l'hypothèse raisonnable d'«incertitude». L'estimation la «plus faible» en ce qui concerne les hommes de race noire dépasse 8 % pour chaque grand groupe d'âge entre 20 et 64 ans. En outre, les estimations AD les plus faibles étaient toujours au-dessus des estimations EPC comparables qui comprenaient des limites d'incertitude (lire Adlakhia et coll., 1991).

Figure 2. Intervalles de confiance du sous-dénombrement des hommes de race noire : estimations AD et EPC de 1990

Figure 2. Undercount Confidence Intervals Black Males: 1990 DA and PES



importe de souligner que les estimations AD présentent moins d'incertitude lorsqu'on mesure les variations de couverture selon l'âge, le sexe et la race. Cette propriété est attribuable au fait que beaucoup d'erreurs dans les estimations sont cohérentes, donc ont tendance à «s'annuler» dans les comparaisons concernant le sexe, la race et le temps. Cet avantage particulier pourrait être exploité en 2000. Ainsi, les ratios hommes-femmes AD (rapport du nombre d'individus de sexe masculin et féminin) sont moins susceptibles de présenter des erreurs que les estimations AD du sous-dénombrement proprement dites.

2.2.2 Forces de l'AD

L'analyse démographique présente certains avantages sur l'approche des enquêtes dont on pourrait tirer parti dans un PMIC général, en 2000. Certains de ces points forts, déjà apparents lors du recensement de 1990, ont gagné en «importance» dans le cadre du Recensement de 2000 pour lequel on vise des économies et la production d'un chiffre de population unique.

1. **Faible coût**--La réduction des coûts figurant parmi les principaux objectifs du Recensement de 2000, le coût relativement faible de l'AD est devenu très attrayant. L'AD est bon marché parce qu'elle recourt abondamment au programme courant d'estimations démographiques du Bureau of the Census. Même si on intensifiait le programme de recherche, l'AD demeure beaucoup moins coûteuse que les méthodes faisant appel aux enquêtes.

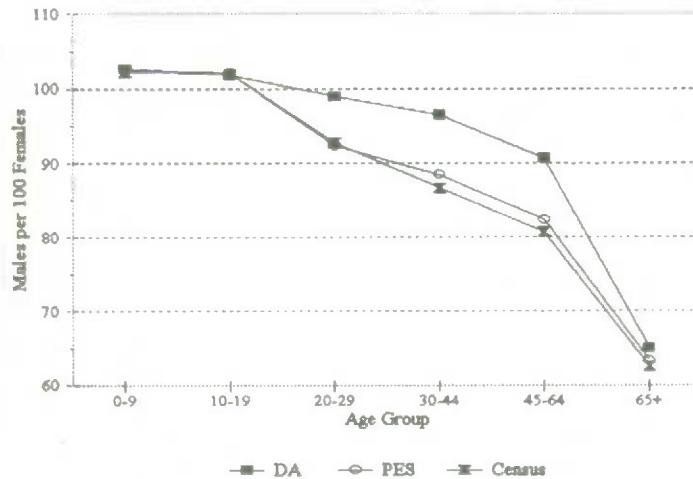
2. **Faisabilité pratique**--L'AD a été testée lors des recensements antérieurs et on n'a cessé d'améliorer les données et les techniques, ainsi que les résultats qui peuvent être évalués. La technique CensusPlus est toujours en rodage et a effectivement soulevé des problèmes imprévus lors de l'essai de 1995. Par ailleurs, l'approche ESD suppose le respect d'un échéancier très serré si on veut obtenir des résultats d'ici le 31 décembre 2000. Les estimations AD indépendantes, reposant sur les dossiers administratifs, pourraient dépanner si les deux méthodes précitées soulèvent des difficultés.

3. **Opportunité**--Puisqu'il n'est pas question d'activités sur le terrain ni de couplage avec le recensement, les estimations AD seront connues avant les estimations de CensusPlus ou ESD en 2000. Tout d'abord, les valeurs de référence indépendantes sur les unités de logement permettraient d'évaluer la complétude du fichier d'adresses principal avant que le Recensement de 2000 proprement dit ne débute. En deuxième lieu, les estimations AD de la population pourraient donner des indices importants sur le sous-dénombrement différentiel lors du dénombrement précédant le «PMIC» (c'est-à-dire en juillet-août 2000). Par exemple, l'indication d'un sous-dénombrement relativement important chez les hommes de race noire selon le ratio hommes-femmes (comme dans les recensements antérieurs) soulignerait qu'il faut poursuivre les activités du PMIC. Bien sûr, les estimations AD seront aussi disponibles pour une évaluation immédiate des estimations par sondage dès que celles-ci seront prêtes (octobre-novembre 2000).

4. **Indépendance**--L'AD repose dans une large mesure sur l'agrégation des dossiers administratifs. Elle crée donc une base indépendante de données permettant de confirmer les estimations par sondage du PMIC. En 1990, les estimations AD indépendantes du sous-dénombrement (1,85 %) ont validé l'estimation globale de l'EPC (1,58 %). Les estimations AD détaillées ont toutefois révélé que l'EPC sous-estimait sensiblement le sous-dénombrement net des hommes de race noire--le problème bien connu du «biais de corrélation». La figure 3 montre de quelle façon le ratio hommes-femmes de l'EPC de 1990 pour les personnes de race noire est beaucoup plus près du ratio hommes-femmes peu plausible du recensement que des ratios de l'AD. Même en tenant compte de l'incertitude des estimations AD et EPC, les ratios hommes-femmes de l'AD restent sensiblement plus élevés que ceux de l'EPC ou du recensement.

Figure 3. Ratios hommes-femmes prévus en 1990 : comparaison de l'AD et de l'EPC au recensement

Figure 3. 1990 Expected Sex Ratios: Comparison of DA and PES to Census



En prévision de l'an 2000, on s'efforce de voir comment intégrer les résultats AD (notamment le ratio hommes-femmes) au PMIC afin d'atténuer le problème le plus possible. Dans ce cas, l'AD servirait clairement deux fins, soit l'évaluation de la couverture et la mesure de la couverture (lire Wolter, 1990 et Bell, 1993 pour plus de détails sur la recherche concernant les ratios hommes-femmes AD dans l'estimation de la couverture).

5. Cohérence interne--La méthode démographique a pour pierre angulaire la cohérence logique et longitudinale des données démographiques sous-jacentes. L'AD suit le processus démographique de l'évolution de la population quand il se produit, à partir de la naissance en augmentant ou en réduisant la taille de la cohorte au moyen de données sur la mortalité et la migration nette. Les estimations obtenues pour l'an 2000 grâce à ce processus seront cohérentes aussi bien sur le plan longitudinal qu'à l'interne. Le couplage chronologique des estimations AD (pour plusieurs recensements) crée une base cohérente permettant d'évaluer la plausibilité des estimations démographiques proprement dites. Parallèlement, les estimations par sondage ne présentent aucune dimension longitudinale et on ne peut en vérifier la cohérence longitudinale et transversale.

Un avantage de l'AD à cet égard est qu'elle donne des estimations détaillées par année d'âge. Les données administratives de l'AD sont à toute fin pratique complètes (pas d'échantillon) et disponibles tous les ans (à savoir, naissances, décès et immigration). Le processus démographique débouche automatiquement sur des estimations détaillées par année d'âge. Les estimations par sondage reposent nécessairement sur un échantillon, ce qui réduit la qualité des estimations par année d'âge. Entre autres utilisations, les données par année d'âge jouent un rôle important dans le programme d'estimations démographiques annuelles du Census Bureau. On pourrait accroître la qualité des données PMIC pour 2000 si on intégrait les estimations AD au processus de mesure de la couverture.

6. Valeurs de référence historiques--Un des principaux objectifs du recensement de 2000 consiste à atténuer la variation du sous-dénombrement. Les estimations AD fournissent la seule série historique cohérente des facteurs de sous-dénombrement précis selon l'âge, le sexe et la race susceptibles d'étayer une éventuelle réduction du sous-dénombrement en 2000, comparativement aux recensements antérieurs. Les estimations par sondage ne présentent pas pareille dimension historique. Par ailleurs, les estimations EPC détaillées de 1990 pour les personnes de race noire comportent des erreurs qui n'autorisent pas une comparaison valable entre les recensements de 1990 et de 2000 (à savoir, les ratios hommes-femmes pour les personnes de race noire ne sont pas plausibles comparativement aux estimations AD).

3. DISCUSSION

En essayant de concevoir un système général de mesure intégrée de la couverture pour le Recensement de 2000, on doit veiller à équilibrer les forces et les faiblesses de l'AD et des techniques d'enquête. Il est évident que l'analyse démographique devrait jouer un rôle important dans l'évaluation du recensement et les activités du PMIC. Les estimations démographiques indépendantes seront disponibles en temps opportun pour que l'on effectue maintes observations sur les tendances liées à la couverture, avant et après l'exercice du PMIC, et cela à un coût relativement faible.

La question qu'on se pose est la suivante : devrait-on aller un pas plus loin et intégrer officiellement l'analyse démographique au processus de mesure de la couverture du PMIC? Plus précisément, peut-on améliorer les estimations du PMIC là où l'AD donne de bons résultats, à l'inverse des estimations par sondage, c'est-à-dire--(1) pour mesurer le sous-dénombrement des hommes de race noire et (2) pour produire des estimations détaillées selon l'âge, le sexe et la race qui présentent les propriétés démographiques de cohérence longitudinale et interne? En intégrant les résultats de l'AD à ceux du PMIC en 2000, on rétrécira l'écart âge-sexe-race entre les estimations AD et les estimations par sondage avant la diffusion du chiffre de population unique du recensement, pas après.

Nous nous efforçons d'élaborer un programme de recherche qui montrera comment intégrer l'AD au PMIC. Ce programme indiquera aussi les travaux nécessaires à une amélioration des estimations AD. L'objectif est clair : exploiter sélectivement et de façon créative les avantages uniques de l'analyse démographique en vue d'améliorer les estimations par sondage du PMIC dont on se servira pour produire le chiffre de population unique définitif du recensement de 2000.

4. BIBLIOGRAPHIE

Adlakha, Arjun, Hogan, Howard, et Robinson, J. Gregory. "A Report on the Internal Consistency of the Post-Enumeration Survey Estimates," *1990 Coverage Studies and Evaluation Memorandum Series S-1*, U.S. Bureau of the Census, 1991.

Bell, William R. 1993. "Using Information from Demographic Analysis in Post-Enumeration Survey Estimation," *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1106-1118

Das Gupta, Prithwis. 1991. DA Evaluation Project D10: "Models for Assessing Errors in Undercount Rates Based on Demographic Analysis." *Preliminary Research and Evaluation Memorandum No. 84*, U.S. Bureau of the Census.

Fernandez, Edward W. 1995. "Using Analytic Techniques to Evaluate the 1990 Census Coverage of Young Hispanics," *Technical Working Paper Series No. 11*, Population Division, U.S. Bureau of the Census.

Himes, Christine L. et Clogg, Clifford C. 1993. "An Overview of Demographic Analysis as a method for Evaluating Census Coverage in the United States," *Population Index*, 58(4): 587-607, hiver 1992.

Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1047-1060.

Mulry, Mary H. et Rajendra P. Singh. 1995. "Development and Evaluation of Census Methodology for 2000 Census," *Proceedings of the International Conference on Survey Measurement and Process Quality*, Bristol, Royaume-Uni, 1er au 4 avril.

Robinson, J. Gregory et Edward L. Kobilarcik. 1995. "Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach." Présenté à l'Annual Meeting of the Population Association of America. San Francisco.

Robinson, J. Gregory. 1994. "Use of Analytic Methods for Coverage Evaluation in the 2000 Census." Présenté à la Population Association of America, Miami, 5 au 7 mai.

Robinson, J.G., Ahmed, B., Das Gupta, P., et Woodrow, K.A. 1993 . "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1061-1071.

Wolter, Kirk M. 1990 "Capture-Recapture Estimation in the Presence of a Known Sex Ratio, *Biometrics*, Vol. 46, P. 157-162.

SESSION 3

MAXIMISATION DES TAUX DE RÉPONSE

LE TAUX DE RÉPONSE ET L'ENQUÊTE SUR LA POPULATION ACTIVE CANADIENNE: FRUIT DU HASARD OU BONNE PLANIFICATION?

Mike Sheridan, Doug Drew et Benoît Allard¹

RÉSUMÉ

Le taux de chômage et les estimations d'emploi selon l'Enquête mensuelle sur la population active canadienne (EPA) sont au nombre des principaux indicateurs économiques clés produits par Statistique Canada. Depuis ses débuts, l'EPA a joui, sur le long terme, d'excellents taux de réponse. Notre communication vise à quantifier les raisons de ce succès. Elle traite des tendances des courbes de réponse et de non-réponse de l'EPA, dont la saisonnalité, et analyse certains des principaux facteurs qui, à l'occasion, ont contribué aux augmentations des taux de non-réponse. En outre, elle présente une évaluation des facteurs et des processus qui contribuent à expliquer le niveau élevé soutenu de réponse. L'analyse de ces facteurs met l'accent sur des questions comme la formation, les méthodes d'interview sur place et par téléphone, la publicité, le fardeau de déclaration, le mode de renouvellement et plusieurs facteurs qui ont des incidences - positives ou négatives - sur les taux de réponse.

MOTS CLÉS: Non-réponse à l'Enquête sur la population active; composantes de la non-réponse; taux de non-réponse; techniques pour réduire la non-réponse; intervieweurs et non-réponse.

1. INTRODUCTION

Le titre de la communication demande, un peu en guise de plaisanterie : «Les bons taux de réponse sont-ils le fruit du hasard ou de l'astuce?» Pour y répondre, nous essayons de faire deux choses. Tout d'abord, nous expliquons un peu la perspective de gestion du problème des taux de non-réponse dans les grands programmes statistiques. Cela s'accomplit sans formules ni notation statistiques. Ensuite, nous disons quelques mots des mesures que Statistique Canada a mises en oeuvre pour réduire la non-réponse et l'erreur systématique inhérente qui l'accompagne habituellement. On peut dire sans trop risquer de se tromper qu'un certain nombre de facteurs incontrôlables définissent les taux finaux de non-réponse de toute enquête. Au contraire de l'approche méthodologique scientifique que proposait Bob Groves ce matin, nous jouons les deux cartes frimées de la non-réponse : le sujet de l'enquête et l'humeur des répondants. Il faut souligner que, comme dans tous les projets dans le monde de la recherche par sondages, la cheville d'attelage dépend, malheureusement, dans une grande mesure de l'ARGENT. C'est l'argent qui, dans une large mesure, impose aux gestionnaires un grand nombre de leurs décisions collectives sur l'importance de la non-réponse qu'une enquête particulière ou un programme d'enquête particulier peut tolérer ou tolérera effectivement.

Nous ne sommes pas non plus téméraires au point de rejeter d'emblée l'idée que le hasard a sa place dans l'équation, mais c'est à contrecœur que nous en venons à la conclusion qu'il n'est pas fondamental. Cela nous rappelle un peu une déclaration de Martin B. Wilk, ancien statisticien en chef du Canada, à qui il arrivait de dire en pontifiant que, parfois, il vaut bien mieux être chanceux qu'astucieux.

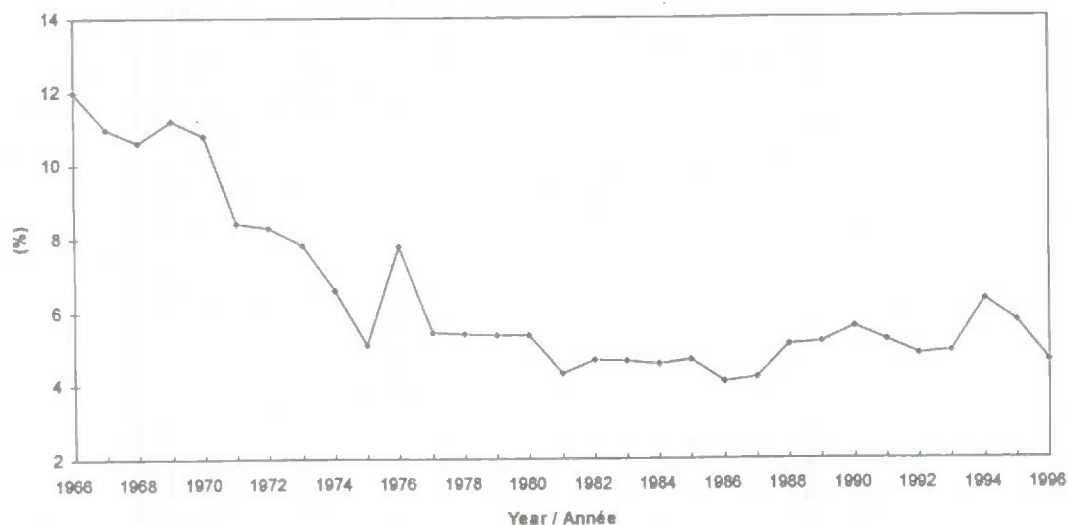
2. LA NON-RÉPONSE DANS L'EPA : DE 1966 À 1996

Pour faire ressortir notre premier point, nous examinons les tendances de la non-réponse globale des 30 dernières années de l'EPA, et analysons certains des facteurs opérationnels influant sur ces tendances.

¹ Mike Sheridan, Directeur, Division des enquêtes des ménages; Doug Drew, Directeur adjoint, Sous-division de l'enquête sur la population active; Benoit Allard, Méthodologiste, Division des méthodes d'enquêtes-ménages, Statistique Canada, Parc Tunney, Ottawa, (Ontario), Canada, K1A 0T6.

Ainsi qu'il ressort de la figure 1, le taux annuel de non-réponse pour l'EPA a diminué considérablement à la fin des années 60 et au début des années 70. Cette diminution semble avoir été associée à un accent accru sur la formation des intervieweurs et le contrôle de la qualité du travail des intervieweurs. Une révision du manuel de l'intervieweur au début des années 70 en a fait un outil plus détaillé, qui comprenait des procédures précises pour les diverses situations surgissant sur le terrain. Ainsi, le manuel donnait des instructions précises sur la façon d'établir les contacts et de faire les rappels. De même, on a institué une procédure spéciale de suivi du lundi dans les enquêtes de juillet et d'août, en vue de rejoindre les ménages qui étaient temporairement absents (surtout en vacances) pendant la semaine de l'enquête.

Figure 1
Overall LFS Nonresponse Rate (Yearly Averages)
Taux de non-réponse global de l'EPA (moyennes annuelles)



L'avènement de l'interview téléphonique a été un autre facteur de réduction de la non-réponse. Au début des années 70, on a commencé à utiliser l'interview téléphonique pour les interviews subséquentes dans les régions urbaines (les interviews initiales se faisaient toujours sur place). En 1975, presque toutes les villes canadiennes avaient été converties à la nouvelle technique. Puisque les appels téléphoniques sont plus faciles et moins coûteux que les interviews sur place, le nombre de tentatives de prise de contact a été considérablement accru, d'où réduction de la composante de la non-réponse attribuable à l'absence de contact.

En 1976, la taille de l'échantillon de l'EPA est passée de 36 400 à 55 700 ménages. Cet ajout de près de 20 000 ménages a nécessité l'embauche et la formation d'un grand nombre de nouveaux intervieweurs. Par conséquent (du moins en apparence), les taux de non-réponse ont été élevés en 1976 (7,6 %), mais sont retombés aux alentours de 5-6 % une fois que l'augmentation de l'échantillon eût été mise en oeuvre, et que les nouveaux intervieweurs eussent pris de l'expérience.

Le taux global de non-réponse a été à son plus bas en 1981 et 1987; le taux annuel a été inférieur à 5 % tout au long de cette période. Entre 1988 et 1993, le taux annuel de non-réponse se situait encore une fois dans la fourchette des 5-6 %. Une autre augmentation de l'échantillon d'environ 16 000 ménages a été mise en oeuvre à la fin de 1989. Cette augmentation pourrait expliquer pourquoi le taux de non-réponse est demeuré un peu plus haut au début des années 90 qu'au milieu des années 80.

À la fin de 1993 et au début de 1994, l'enquête a été convertie de l'interview papier et crayon (IPC) à l'interview assistée par ordinateur (IAO). Ce changement a déclenché une augmentation du taux de non-réponse. Les versions initiales du logiciel IAO d'applications et de gestion des cas ont ralenti les interviews, de sorte qu'il restait moins de temps pour des tentatives multiples de prise de contact - d'où augmentation de la non-réponse imputable à l'absence de contact. Divers problèmes techniques ont aussi

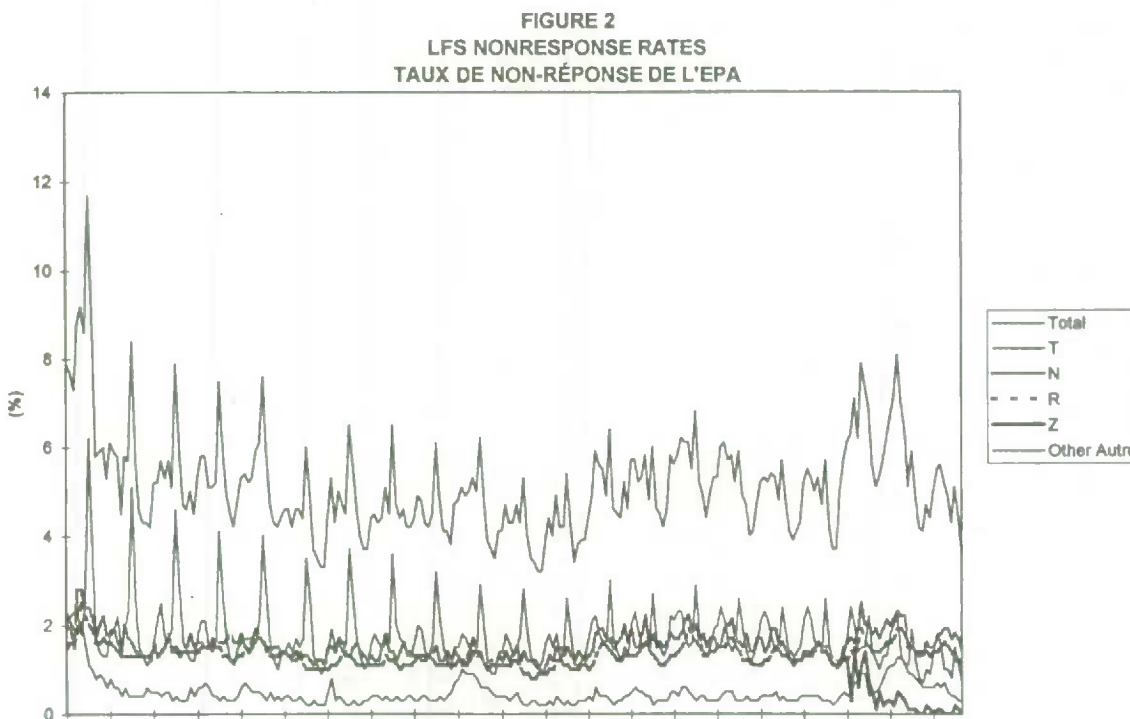
provoqué la perte de données sur les machines des intervieweurs ou en cours de transmission. Ces deux problèmes ont été graduellement éliminés avec l'amélioration des applications et des processus d'IAO, si bien que le taux de réponse s'est amélioré jusqu'au milieu de 1994.

À la fin de 1994 et au début de 1995, un nouveau plan de sondage a été graduellement intégré, ce qui a encore une fois propulsé la non-réponse. De nombreux nouveaux intervieweurs ont été alors embauchés, et les charges de travail ont augmenté en raison des activités de listage nécessitées par l'introduction d'un nouvel échantillon. De même, pour ne pas alourdir excessivement le fardeau du répondant, il a été décidé qu'un ménage échantillonné selon le nouveau plan n'était pas tenu de répondre à l'enquête s'il avait fait partie de l'échantillon au cours des deux années précédentes. Ces cas de non-réponse imputables au «chevauchement d'échantillons» représentaient 0,6 % de l'ensemble des ménages échantillonnés à la fin de l'intégration graduelle du nouveau plan de sondage en mars 1995.

Dans les mois qui ont suivi, le taux de non-réponse a chuté rapidement, retombant au faible niveau d'avant l'introduction de l'IAO. Les taux de réponse ont réagi aux deux grands changements mis en oeuvre depuis la fin de 1993 (la conversion à l'IAO et la refonte du plan de sondage), mais temporairement seulement; le taux moyen de non-réponse en 1996 était de 4,6 %, soit le plus bas depuis 1987.

2.1 Les composantes de la non-réponse de l'EPA

La figure 2 présente le taux mensuel de non-réponse de l'EPA depuis 1976. Les courbes inférieures du graphique montrent la répartition du taux global de non-réponse en cinq catégories :



1. *Temporairement absent (T)* : Le ménage est absent de la maison pendant la période d'interview (en vacances, par exemple). L'intervieweur en a été informé lors d'une interview précédente, ou l'a appris par un voisin.
2. *Personne à la maison (N)* : L'intervieweur n'a pu établir de contact.
3. *Refus (R)* : Le ménage choisi a refusé de participer à l'enquête.
4. *Problème technique (Z)* : Cela comprend les problèmes de transmission liés à l'IAO, qui sont mentionnés plus haut, et les cas perdus à cause de pannes d'ordinateur portable.
5. *Autre* : Cela comprend toutes les autres raisons de la non-réponse : mauvais temps, circonstances hors de l'ordinaire au sein du ménage (comme un décès ou une maladie), barrière linguistique, manque d'intervieweur, etc.

Ces descriptions détaillées et scientifiques se ramènent essentiellement à trois sources de non-réponse :

1. nous ne pouvons les rejoindre : soit qu'ils ne sont pas à la maison, soit que nous les rattrapons toujours,
2. nous les trouvons et ils refusent de répondre,
3. avec la nouvelle technologie, nous les trouvons, ils nous répondent, puis nous ne pouvons sortir du cyberspace.

Comme il ressort de la figure 2, le plus grand facteur de la non-réponse mensuelle pour l'enquête pour une période prolongée a été les absences temporaires, suivies des cas où il n'y a personne à la maison, puis des refus, et enfin des problèmes techniques. Abstraction faite des facteurs saisonniers - et nous y reviendrons - la figure 2 permet de se demander pourquoi les absences temporaires régressent sur une période de vingt ans. Nous soupçonnons que la découverte de l'explication nous lancerait dans un long et intéressant voyage. Notre hypothèse comprend à la fois des facteurs économiques et sociaux, et nous gardons cela pour une autre communication.

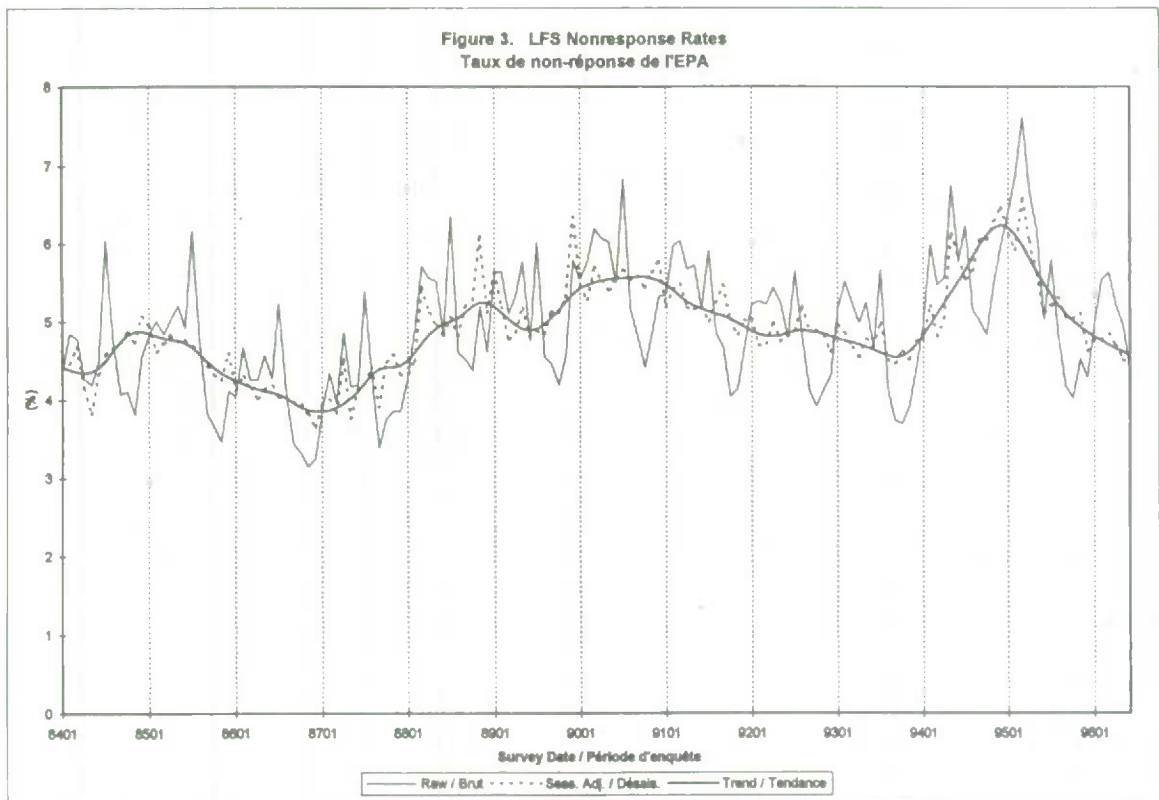
Au cours des 20 dernières années, il y a eu une légère diminution du nombre de refus, qui s'inscrit véritablement dans une tendance à la baisse à long terme, interrompue seulement par certaines petites fluctuations saisonnières et mensuelles. La décision de gestion est, bien sûr, de savoir combien d'autre d'argent il faut dépenser pour abaisser encore davantage la non-réponse...

2.2 Saisonnalité

Comme nous l'avons mentionné plus tôt, la non-réponse a aussi une composante saisonnière. La figure 3 indique la non-réponse, brute et désaisonnalisée, avec courbe de tendance intégrée dans le graphique. Le taux global de non-réponse suit une tendance saisonnière manifeste, caractérisée par une pointe en juillet. Cette pointe, qui a été très élevée à la fin des années 70, a diminué graduellement pendant les années 80. Entre-temps, la non-réponse a augmenté dans les mois d'hiver. Par conséquent, le régime saisonnier observé dans les années 90 est caractérisé par des creux marqués au début des mois d'automne (septembre et octobre) plutôt que par les pointes élevées de juillet de la fin des années 70 et du début des années 80 (la pointe de juillet apparaît encore dans les années 90, mais elle a beaucoup moins d'influence qu'auparavant). En revenant à la figure 2, on voit clairement que la saisonnalité du taux global de non-réponse est surtout fonction de la composante «Temporairement absent» (T). Le changement de saisonnalité est aussi très apparent dans la composante «Temporairement absent».

Le taux de refus (R) contribue également dans une certaine mesure à la saisonnalité du taux global de non-réponse, en atteignant son point culminant en mai et son point creux à l'automne. Le taux de réponse est élevé en avril et mai, grâce à deux grandes enquêtes supplémentaires (l'enquête sur les finances des consommateurs et l'enquête sur l'équipement ménager), qui sont administrées pendant ces mois à un gros sous-échantillon de l'EPA. De même, les Canadiens doivent remplir leur déclaration de revenus pour la fin d'avril de chaque année, et cet exercice peut avoir un effet sur l'humeur du public, qui perçoit une autre corvée gouvernementale, si bien que certains ménages sont peut-être moins portés que d'autres à répondre. La tendance traduit une augmentation des taux de non-réponse sur la période 1987-1991, puis de nouveau entre 1994 et 1995. Depuis 1995, le taux a chuté, revenant à peu près au même niveau que 12 ans plus tôt, en 1984. C'est peut-être là qu'on pourrait dire quelques mots sur le «hasard» et le «facteur des cartes

frimées» : nos répondants et le grand public. L'augmentation des taux de réponse pendant la période du milieu de 1987 au premier trimestre de 1991 est le résultat d'un certain nombre de facteurs, dont certains pourraient comprendre les effets déstabilisateurs possibles de l'introduction d'une augmentation d'échantillon, les incidences d'une conjoncture difficile, reflétée par la récession, ou peut-être tout simplement l'humeur du public. On se demande quelles peuvent être l'incidence de la popularité d'un gouvernement et sa relation possible avec le taux de réponse. Avec certaines analyses de données que nous avons faites, on pourrait faire valoir que l'humeur du public a des incidences sur les relations avec les répondants et surtout lorsque l'interlocuteur est un service du gouvernement fédéral comme Statistique Canada, qui joue un rôle de mandataire de la Couronne. Pendant cette période d'accroissement de la non-réponse, les cotes d'approbation, mesurées par diverses sociétés de sondages du secteur privé, ont diminué, ce qui a créé une relation réciproque entre les taux de non-réponse de l'EPA (augmentation) et la cote d'approbation du gouvernement (diminution). Le fléchissement des cotes d'approbation et peut-être le lien avec notre non-réponse accrue ont pu refléter une certaine insatisfaction des répondants à l'égard de la politique gouvernementale. Par ailleurs, il se peut que ce soit seulement l'économie. Ou encore, cela a pu être des facteurs opérationnels, comme l'effet du recrutement d'un grand nombre de nouveaux intervieweurs coïncidant avec un grand accroissement de l'échantillon. Nous laissons à d'autres le soin de faire l'analyse de régression et la modélisation, et laissons là les conjectures.



2.3 Renouvellement

Selon le plan de renouvellement de l'EPA, les ménages échantillonnés sont renouvelés chaque mois par sixième. L'un des phénomènes très étudiés de l'enquête est les taux de non-réponse sur les six mois que les ménages demeurent dans l'enquête. La figure 4 donne le taux de non-réponse par nombre de mois que les ménages ont été dans l'enquête. Le tableau montre une période de 12 mois, en 1995 et 1996, qui est représentative de la tendance typique observée. La tendance est à l'augmentation de la non-réponse dans le premier mois que les ménages sont dans l'échantillon, conséquence surtout de l'augmentation des taux

d'absence de contact. La non-réponse chute dans les mois suivants. Cela s'explique par plusieurs raisons. Dès lors qu'ils ont pris contact avec un ménage, les intervieweurs connaissent les meilleures heures pour rappeler les mois suivants. Un autre facteur est que la première interview se fait généralement sur place, de sorte qu'il y a une limite pratique au nombre de tentatives de prise de contact possibles dans la brève période de collecte.

FIGURE 4
NONRESPONSE RATES BY NUMBER OF MONTH IN THE SURVEY FOR CANADA -
TOTAL AND BY COMPONENT /TAUX DE NON-RÉPONSE PAR NOMBRE DE MOIS
DANS L'ENQUÊTE POUR LE CANADA AU TOTAL ET PAR COMPOSANTE
MEAN FOR SURVEYS 9508 TO 9607/MOYENNE DES ENQUÊTES 9508 À 9607

Number of month/ Nombre de mois	1	2	3	4	5	6	Total
Total	7.53	4.76	4.33	4.23	4.19	3.93	4.82
Temporarily absent/ Temporairement absent	1.84	1.21	1.06	0.99	0.96	0.84	1.15
No one at home/ Personne à la maison	2.96	1.63	1.45	1.31	1.21	1.03	1.60
Refusal/Refus	1.72	1.28	1.26	1.34	1.38	1.39	1.39
Other/Autre	1.01	0.64	0.56	0.59	0.64	0.67	0.68

Un des grands facteurs du recul de la non-réponse d'ensemble a été l'introduction des interviews téléphoniques pour les mois 2 à 6. Cette procédure - qui consiste à téléphoner après une interview initiale sur place - s'appelle l'interview *de mise en confiance*. Cette approche a peu changé depuis son lancement au début des années 70. L'incidence du téléphone a été de ramener le taux de non-réponse de la fourchette des 10-12 % du début des années 70 à celle des 5-6 % d'aujourd'hui.

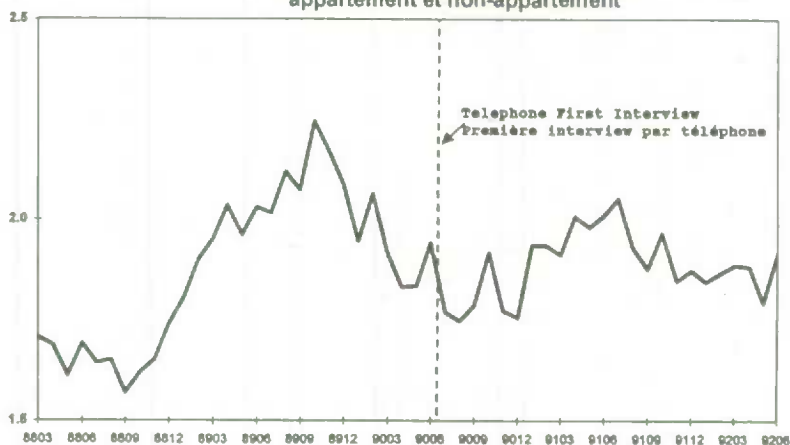
L'approche de l'interview de mise en confiance était basée sur le sentiment, établi depuis longtemps, de confort et de bien-être que procure le fait que la première des six interviews se fait sur place. De fait, le sentiment de confort se traduit non seulement dans les taux de réponse, mais encore, jusqu'à un certain point, dans la conviction que, d'une certaine façon, le premier contact personnel améliore la qualité d'ensemble des données. Cette conviction existe, même s'il n'y a pas beaucoup de données empiriques pour appuyer la thèse de la qualité des données. Nous avons examiné à fond ce phénomène pour voir si nous avons affaire à une réalité ou à un mythe pour ce qui concerne la question des interviews sur place du premier mois et leur relation et leur effet généralement tenu pour positif sur les taux de réponse des mois suivants.

En premier lieu, il a été reconnu que le téléphone est devenu un élément «clé» du dénouement de l'énigme de la façon de communiquer avec des personnes très mobiles, rarement à la maison, qui ont tendance à être jeunes et célibataires et à vivre dans des tours d'habitation. L'EPA a décidé en 1990 d'instituer les interviews par téléphone des nouveaux répondants dans la partie des tours d'habitation de l'échantillon de l'EPA, notamment pour contourner le problème des mécanismes de sécurité liés à ces immeubles, ainsi que pour contribuer tout simplement à surmonter la difficulté de trouver quelqu'un à la maison pendant une visite sur place dans cette catégorie de logements. La figure 5 illustre les incidences de la décision de faire par téléphone les interviews du premier mois pour les tours d'habitation comprises dans l'échantillon. Elle montre les taux différentiels de non-réponse entre l'échantillon d'appartements avant et après la mise en oeuvre de la première interview par téléphone, et il n'y a pas beaucoup de différence dans la comparaison.

La non-réponse dans les tours d'habitation qui font partie de l'échantillon est demeurée environ deux fois plus forte que dans les autres genres de logements, et l'introduction de la première interview téléphonique n'a ni amélioré ni empiré considérablement la situation. Les plus fortes incidences de l'utilisation de cette approche dans les logements de l'échantillon qui font partie de tours d'habitation sont la réduction des coûts des interviews pour les logements échantillonnés.

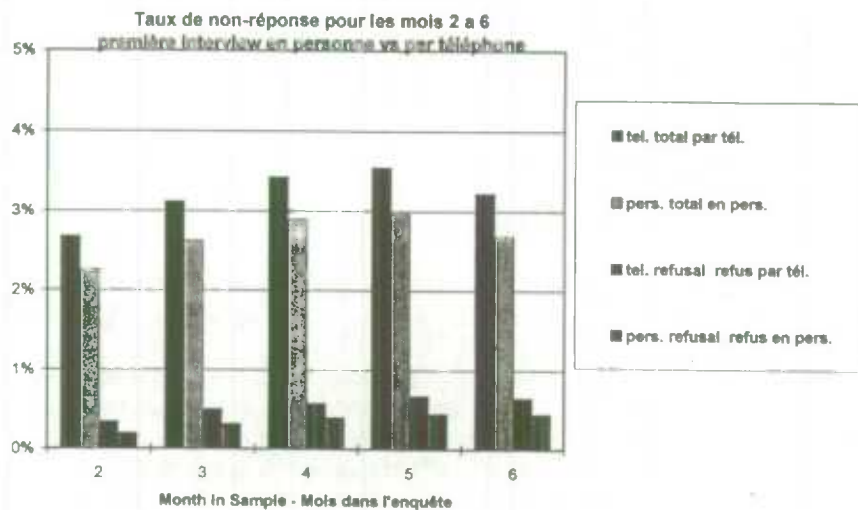
Figure 5
Ratio of first month nonresponse rates:
apartment vs non-apartment

Rapport de taux de non-réponse pour le premier mois:
appartement et non-appartement



À l'heure actuelle, pour l'ensemble de l'échantillon de l'EPA, environ 75 % des interviews du premier mois se font par téléphone, ce qui laisse environ 25 % d'interviews des nouveaux répondants. Le nombre de créations par téléphone est en progression depuis quelques années. Le contact téléphonique du premier mois est essentiellement un moyen de suivi pour les cas où il n'y a pas eu de contact. Les intervieweurs ont pour instructions de faire tout leur possible pour effectuer une interview sur place, mais ils ont la permission d'utiliser le téléphone dans les tentatives suivantes s'ils n'arrivent pas à prendre contact.

Figure 6
Nonresponse rates months 2 to 6:
personal vs telephone first interview



La figure 6 présente la relation entre les taux de non-réponse dans les mois 2 à 6, selon que l'interview était sur place ou par téléphone dans le premier mois. Ces résultats supposent, naturellement, qu'il y a un répondant dans le premier mois. Et puisque, dans le premier mois, on utilise le téléphone surtout pour faire le suivi des ménages avec lesquels il n'y a pas encore eu de contact après une visite sur place, le groupe «téléphone» a une propension plus ou moins grande à la non-réponse - du moins de la variété de l'absence de contact - dès le départ. Abstraction faite de cette mise en garde, il y a quelques tendances immédiatement manifestes. La première est que la non-réponse totale pour les ménages où la première interview a lieu par téléphone est toujours plus haut dans les autres mois de l'échantillon. L'autre évidence qui ressort est que les taux de refus, dans le premier mois ne sont que légèrement plus hauts pour les interviews téléphoniques que pour les interviews sur place.

Donc, même si l'on fait abstraction des différences de population, et même si l'on accepte l'idée que l'interview du premier mois contribue à l'amélioration des taux de réponse dans les mois suivants, cela pose encore la question, dans la perspective de gestion et dans une perspective d'efficacité, de savoir si ces différences plutôt faibles de performance valent l'investissement et le coût vraiment très élevés des interviews personnelles pour le premier mois. La question étant posée, nous voudrions faire une analyse plus détaillée et plus rigoureuse de l'interview sur place du premier mois. La plus grande des questions auxquelles il faudrait répondre, bien sûr, toucherait les incidences sur la qualité des données. Une autre complication qui s'ajoute au tableau est la nécessité de faire une évaluation des incidences de l'interview sur place sur les taux de réponse pour les autres enquêtes qui utilisent la capacité d'enquêtes supplémentaires de l'EPA. En outre, des échantillons des ménages de l'EPA, après leurs six mois de participation à l'EPA, sont parfois utilisés comme échantillon initial d'enquêtes longitudinales. Donc, s'il ne semble pas y avoir d'incidences considérables sur les taux de réponse de l'EPA, d'autres utilisateurs de la capacité de l'EPA ont peut-être moins de succès à fidéliser les répondants sans l'avantage d'une première interview de création d'empathie «théorique».

L'autre facteur des taux de réponse de l'EPA est une chose dont ne profitent pas les recherches ponctuelles et privées, et c'est la publicité qui est faite à l'enquête chaque mois auprès des médias et des décideurs. Cette exposition, conjuguée aux dispositions du chapitre 19 des Lois du Canada, présente certains avantages du côté «réponse» de l'équation. Mis à part ces avantages manifestes, ceux qui travaillent avec le programme de l'EPA feraient valoir que plusieurs éléments clés sont critiques pour la définition du succès de cette enquête et des enquêtes supplémentaires connexes. De fait, on pourrait dire que ce sont des éléments génériques applicables à toute la recherche par enquêtes.

1. Recrutement et embauche. On ne saurait surévaluer l'importance de trouver les bonnes personnes pour le travail. Dans le cas de l'EPA, ces personnes ont aussi un engagement envers le programme.
2. Formation. Il ne sert absolument à rien de même songer à effectuer un travail décent de recherche par enquêtes sans un vigoureux investissement dans la formation. Dans le programme de l'EPA, l'accent sur la formation est double. Le premier volet est l'apprentissage de ce qu'il faut pour effectuer le travail, et le second est le maintien et l'amélioration de ces techniques.
3. Culture. La non-réponse n'est pas acceptable, ni pour les régions ni pour l'administration centrale. Il peut sembler y avoir parfois un conflit entre le niveau élevé des taux de réponse et l'impératif de qualité des données, mais à long terme l'équilibre s'est établi.
4. Suivi à faire en cas de refus. Un intervieweur principal effectue un suivi, le directeur régional envoie une lettre en cas de refus et il y a une nouvelle visite le mois suivant.
5. Outils et appui de l'intervieweur. Ils comprennent une lettre d'avis de visite, une lettre de présentation aux répondants, un numéro de téléphone pour les répondants difficiles à rejoindre. En outre, chaque tâche d'intervieweur donne lieu à un rapport de validation qui renseigne sur les écarts des réponses révisées et qui est analysé chaque mois.
6. Programmes d'observation et de validation. Au-delà de la formation annuelle, chaque intervieweur est observé par l'intervieweur principal et il y a une séance de revue et d'évaluation avec l'intervieweur. Chaque mois, sauf en avril et en décembre, la tâche de 10 % des intervieweurs de l'EPA fait l'objet d'une vérification. On reprend contact avec un échantillon aléatoire de sept ménages de chaque tâche choisie. Le personnel du bureau régional pose un certain nombre de questions aux répondants pour confirmer que l'intervieweur a pris contact avec le ménage et l'a interviewé.

7. L'effet recensement. J'ai de sérieuses réserves au sujet du recensement comme élément qui améliore les taux de réponse. Cependant, cette école a ses ardens défenseurs. Ils attribuent une certaine prime à la publicité du recensement qui incite les gens à répondre ainsi qu'à la publicité entourant les poursuites dans le cadre du recensement, et affirment que ces publicités sont un gain fortuit pour l'EPA.
8. Empathie. Bien que nous ayons démontré que la création d'empathie ne fait pas une si grosse différence dans les taux de réponse, il reste qu'elle joue un rôle intangible important. Le facteur clé dans le processus global est peut-être l'intervieweur même. Certaines personnes savent vraiment s'y prendre, aiment ce travail et restent intervieweurs pendant de nombreuses années, alors que d'autres aiment moins cela et ont tendance à abandonner.

4. CONCLUSION

1. D'abord et avant tout, les taux de non-réponse de l'EPA comme tels ne semblent pas poser de problème.
2. Le facteur du public et sa réaction - celle des médias aussi, d'ailleurs - ont exercé et continueront d'exercer sur les taux de non-réponse des influences qui se prêtent mal à une formulation scientifique précise et qu'il sera difficile, sinon impossible, de prédire et encore plus de contrôler.
3. Les incidences des interviews sur place dans le premier mois, et les effets à long terme sur les taux de réponse méritent une autre analyse serrée dans une perspective d'analyse coûts-avantages.
4. L'EPA a un modèle très simpliste pour huit choses que nous faisons bien dans la collecte. Nous supposons que toutes sont basées sur l'engagement envers le programme des enquêtes-ménages et Statistique Canada. Selon le modèle, tant que nous aurons leur approbation positive et répondrons à leurs attentes, nos intervieweurs continueront d'assurer la qualité des données et les taux de réponse.

5. BIBLIOGRAPHIE

Kumar, S. et Bérard, H. *Nonresponse Rates and Trends in Household Surveys*. Methodology Working Paper SSMD-94-002. Statistique Canada.

Statistique Canada. *EPA-100, Manuel de l'intervieweur*.

MÉTHODES DE RECHANGE LORS DE L'ÉLABORATION DE STRATÉGIES DE COMMUNICATIONS: AUGMENTER LES TAUX DE RÉPONSE OU LES TAUX DE NON-RÉPONSE

Scott D. Nowlan¹

RÉSUMÉ

Le présent document de travail a pour principal objectif de déterminer si l'élaboration d'une stratégie de communications peut conduire à l'augmentation des taux de réponse des participants éventuels à une enquête. Une telle action ne risque-t-elle pas d'engendrer un biais des répondants ou une erreur de non-réponse une fois le contact établi avec le répondant?

Nous poursuivons un certain nombre d'objectifs. D'abord, nous cherchons à expliquer ce qu'une stratégie de communications signifie pour une enquête. Nous présumons au départ qu'à l'étape de la planification de n'importe quelle méthode de collecte de données, y compris la mise au point d'une stratégie et d'une base de sondage, le chercheur devrait songer à procéder à une «prise de contact préalable» avec les répondants afin de les encourager à participer à l'enquête. Même si elle ne constitue pas une solution universelle, la recherche ciblée et plus spécifique, où la base de sondage est bien définie et bien accessible, constitue le type de recherche qui peut se prêter à une telle démarche. Les types d'enquêtes où les répondants sont sélectionnés à partir de données administratives peuvent être cités en exemple. Les enquêtes dont il sera question dans le présent document comprennent l'enquête de 1995 sur les personnes touchant des prestations d'invalidité du Régime de pension du Canada (conduite par Statistique Canada pour Développement des ressources humaines Canada) et l'enquête nationale de 1995 sur les détenus (conduite par Price Waterhouse pour Services correctionnels Canada), qui ont toutes deux bénéficié d'une stratégie de communications élaborée. Nous examinons également les composantes d'une stratégie fonctionnelle et efficace, tout en examinant les effets d'une telle stratégie sur l'augmentation des taux de réponse.

MOTS CLÉS: Communications; non-réponse; taux de réponse.

1. INTRODUCTION

1.1 Observations préliminaires

La conduite d'une enquête peut représenter une tâche ardue, même dans des conditions idéales. Les chercheurs connaissent et savent apprécier les conséquences méthodologiques et analytiques des défis qu'ils ont à relever. Par ailleurs, comme si cela ne suffisait pas, l'arrivée sur le marché des enquêtes d'un nombre accru de participants n'a rien fait pour aider les chercheurs à gagner la confiance des répondants éventuels et surtout, à obtenir leur participation. Cette évolution du marché a eu une incidence directe sur le fardeau des répondants. Les défis posés par cette situation constituent le principal objet de la présente discussion.

En notre qualité de chercheurs, nous devons aller au-delà de ce que j'appellerais les questions de «méthodologie pure», qui intéressent les étapes de la conception des enquêtes et de la vérification et du traitement des données, pour porter notre attention sur les méthodes «non traditionnelles». Compte tenu du milieu concurrentiel dans lequel nous nous trouvons - un milieu où les chercheurs universitaires ainsi que ceux des secteurs publics et privés se disputent l'accès aux répondants - le présent débat est non seulement important, mais même essentiel si nous voulons pouvoir compter sur une source de données primaires provenant du même bassin de répondants.

¹ Scott D. Nowlan, Price Waterhouse, 180, rue Elgin, pièce 1100, Ottawa (Ontario), Canada K2P 2K3.

La non-réponse est un autre problème qui hante continuellement les chercheurs responsables des enquêtes. Le degré de non-réponse a une incidence directe sur l'aptitude du spécialiste de la méthodologie ou de l'analyse à mesurer exactement les résultats, à examiner les faits et à tirer ses conclusions. Ce problème est une raison supplémentaire pour examiner les méthodes de rechange envisageables.

1.2 Sommaire

Le présent document a pour objectif de stimuler les discussions et, peut-être, de susciter un débat sur l'utilisation des méthodes d'enquêtes non traditionnelles. Plus précisément, nous voulons examiner quelle pourrait être l'incidence, sur les taux de réponse et sur les risques de non-réponse, de la décision de «recontacter» les répondants. Nous n'avons pas cherché à rassembler des données afin de prouver une hypothèse, mais plutôt à examiner une démarche utilisée pour mettre au point une stratégie détaillée de communications pour les répondants. Nous examinerons le cadre d'une telle stratégie ainsi que les éléments détaillés de cette démarche. Deux études de cas seront présentées pour illustrer notre propos. Le succès relatif de cette stratégie est toujours sujet à discussion. Néanmoins, nous avons également examiné certains des avantages et des inconvénients prévisibles de cette démarche, et nous avons proposé des moyens d'améliorer la stratégie. Cette stratégie n'a pas fait l'objet d'une évaluation quantitative. Nous nous sommes plutôt appuyés sur des études semblables réalisées par le passé (Dillman, 1978; deVaus, 1991) afin d'examiner des méthodes de rechange qui cherchent à accroître les taux de réponse, à réduire les taux de non-réponse et à augmenter la validité des données que nous recueillons.

2. QU'EST-CE QU'UNE STRATÉGIE DE COMMUNICATIONS?²

2.1 Cadre de l'étude

Toute stratégie de communications doit poursuivre des objectifs précis. Un des plus importants consiste à accroître le taux de réponse obtenu dans une enquête. Le problème consiste toutefois à savoir dans quelle mesure une telle stratégie risque d'accroître le biais dans les réponses. Nous reviendrons plus tard sur cette question. Toutefois, l'examen du problème du biais dans les réponses ne sera abordé ici que superficiellement.

Le temps est une variable importante à prendre en compte dans l'examen de notre stratégie. Il faut du temps pour élaborer et mettre en oeuvre un plan de communications sérieux. Dans un contexte où beaucoup de nos clients souhaitent obtenir rapidement les informations dont ils ont besoin, on passe outre à l'application de cette stratégie. Pourtant, les résultats prévisibles de sa mise en oeuvre sont suffisamment prometteurs pour justifier un tel investissement de ressources. Si cette possibilité est examinée avec les clients et les collaborateurs à l'étape initiale de la conception d'un produit, elle pourra devenir partie intégrante du projet. Cette démarche sera de toute manière préférable à un ajout de dernière minute. Le temps constitue toujours un défi, mais la mise en oeuvre d'une telle stratégie pourrait profiter à l'équipe de chercheurs tout au long de la conduite de leur projet.

L'élaboration d'une stratégie de communications s'appuie sur un certain nombre de principes fondamentaux. Il est important d'aborder ces principes séparément pour traiter correctement du problème de la non-réponse. Dans beaucoup de cas, les répondants pourront refuser de participer simplement parce qu'ils douteront de l'utilité de parachever l'enquête (Dillman, 1978). Même si l'importance de ces «sous-

² L'article de Dillman (1978) constitue une autre source importante d'informations sur l'élaboration d'une stratégie de «contact préalable». Son examen d'une «méthode de conception intégrée» (*Total Design Method* ou TDM) donne également un aperçu intéressant des raisons pour lesquelles l'élaboration d'une stratégie de communications peut être utile. Cette source de renseignements supplémentaires, que nous n'avons découverte que récemment, s'est avérée précieuse pour nous.

éléments» peut paraître évidente aux chercheurs ou aux spécialistes de la méthodologie, chacun d'eux devra être examiné et conçu séparément puisque chacun poursuit des objectifs particuliers. Toutefois, lorsqu'ils sont intégrés, ces sous-éléments doivent permettre d'atteindre les principaux objectifs qui sont d'accroître le degré de réponse, d'informer les répondants et d'obtenir leur collaboration. La stratégie de communications que nous proposons comporte huit de ces sous-éléments : l'événement, l'objectif, les attentes, la confidentialité, l'encouragement, le contact, la transmission de l'information et l'appréciation.

2.2 L'événement

L'enquête que l'on se propose de réaliser devrait être décrite clairement dans l'introduction. On fournira aux répondants éventuels des informations sur le questionnaire ou sur le chargé d'enquête. Cette démarche pourrait être considérée comme un moyen d'«avertir» les répondants qu'ils seront contactés par courrier ou par téléphone. Il convient de mentionner que cet avertissement pourrait devenir le premier «signal d'alarme» entraînant la non-réponse. Un des objectifs des autres éléments de ce contact sera donc de répondre aux préoccupations immédiates ou aux questions qui pourraient être soulevées dès le début. Si le plan de communications est convenablement élaboré, on peut espérer que les autres composantes que nous aborderons plus loin contribueront à éliminer cette «menace» possible au parachèvement de l'enquête.

2.3 L'objectif

Il est vital de bien articuler l'objectif poursuivi pour le bénéfice des répondants éventuels. Il convient en effet d'établir des rapports de «confiance mutuelle» entre chercheurs et répondants. Le scepticisme et l'incompréhension peuvent conduire au refus de collaborer ou à la non-réponse; la communication des objectifs déclarés de la recherche contribuera à atténuer une telle méfiance. Il convient par ailleurs de faire en sorte que tous les intervenants participent à la détermination des objectifs. Nous entendons par intervenants le «client» (ou propriétaire des données de la recherche) et le «chercheur» (ou producteur des données de recherche). On devrait indiquer à quoi serviront les données recueillies, c'est-à-dire, quels sont les objectifs de la recherche.

2.4 Les attentes

Le plan de communications devrait indiquer comment la recherche sera réalisée. Il importe que le répondant soit mis au fait des aspects logistiques de la prise de contact et du parachèvement de l'enquête. À défaut de cela, les répondants risquent de ne pas voir l'importance de compléter l'enquête. En se penchant sur les attentes, on a dans certains cas réussi à accroître le taux de réponse en permettant aux chercheurs d'être plus sensibles aux obstacles à surmonter. On peut mentionner à titre d'exemple la prise en compte de la nécessité, pour certains répondants, d'avoir accès à des appareils de télécommunications pour les sourds (ATS) ou du fait que certains répondants n'ont pas accès à un téléphone. Des méthodes de collecte des données de rechange peuvent être planifiées à l'avance pour tenir compte de ces problèmes et réduire ainsi l'incidence d'un biais de non-réponse ou de réponse. Ici encore, la diffusion de ces informations contribue à mettre les répondants éventuels en confiance. En évitant de donner aux répondants l'impression qu'on leur cache quelque chose, on fera en sorte qu'ils soient mieux disposés à participer et à répondre à nos questionnaires.

2.5 La confidentialité

On ne saurait trop insister sur cet aspect essentiel de toute stratégie de communications efficace. Les chercheurs sont tous conscients du cynisme grandissant manifesté à l'égard du respect de la vie privée. Les gens craignent que les informations de nature personnelle ou commerciale les concernant soient partagées avec d'autres ou utilisées contre eux.³ Les normes d'éthique de notre industrie interdisent de telles pratiques. Toutefois, la simple existence de ces normes ne suffira pas en soi à convaincre le public de participer à nos projets de recherche. Je crois qu'il ne suffit pas de simplement adhérer au concept de la confidentialité, mais qu'il faut constamment en rappeler l'existence à nos répondants éventuels. Des répondants convaincus que les renseignements qui les concernent ne risquent pas d'être mal utilisés participeront plus volontier à nos recherches. En outre, je crois qu'il conviendrait d'aborder les questions de la confidentialité et de l'anonymat à plus d'une occasion. En plus d'être mentionnées dans le cadre de la stratégie de communications, ces questions devraient être abordées de nouveau une fois le contact établi en vue de l'administration d'un questionnaire. On ne saurait sous-estimer l'importance d'une telle démarche. Les personnes qui comprennent bien la notion de confidentialité et qui, en plus, croient à notre sincérité seront plus enclines à participer à nos travaux. Tout indice tendant à prouver le contraire travaillera en notre défaveur et minera la confiance des répondants.

2.6 L'encouragement

Il importe d'insister sur le fait que les chercheurs dépendent de la participation des répondants pour assurer la bonne exécution de leur projet de recherche. Il convient donc, à cette fin, d'encourager la participation des répondants. On peut, par exemple, rappeler aux répondants que leurs opinions et leur comportement ont une incidence importante sur la réussite du projet (voir Dillman); répéter comment leurs réponses influenceront sur les données et sur les résultats de la recherche (p. ex., l'évaluation d'un programme ou un sondage auprès des employés); et leur rappeler qu'ils ont été choisis au hasard et que leur opinion est représentative de celles d'autres personnes (p. ex., données pondérées après coup). Il s'agit de bien montrer aux répondants l'importance de leur participation. L'appel à l'altruisme des gens ne sera pas toujours couronné de succès. On y parviendra plus sûrement en leur expliquant leur rôle et en leur démontrant son importance.

2.7 Le contact

La mention du nom d'une personne à contacter ou d'une adresse ou d'un numéro de téléphone est une autre bonne façon de démontrer le sérieux de notre recherche. Il convient toutefois de rappeler que cette composante de la stratégie de communications risque de compliquer la collecte des données. Par exemple, les chercheurs doivent se tenir prêts à répondre aux demandes d'information. Il convient par ailleurs de noter que cette prise de contact ne servira pas nécessairement à établir ou à augmenter le climat de confiance. Beaucoup de répondants utiliseront cette ressource simplement pour s'assurer de la légitimité du projet de recherche.

Cette précaution n'en présente pas moins certains avantages. Certains répondants l'utiliseront pour organiser la collecte des données (p. ex., pour confirmer ou modifier la date de collecte des données ou la méthode employée), pour obtenir des éclaircissements sur les objectifs déclarés et pour poser des questions précises. Les réponses éventuelles à de telles demandes devraient être examinées pendant l'élaboration de la stratégie de communications. La façon dont on répondra à ces questions pourrait influencer sur la décision des participants éventuels de collaborer à l'enquête. Ce service risque toutefois d'être utilisé pour refuser

³ Price Waterhouse a réalisé de nombreuses entrevues de groupes et recherches qualitatives à ce propos, pour le bénéfice de Statistique Canada. Ces essais de questionnaires auprès de répondants éventuels donnent à conclure à l'existence d'une confusion persistante entre la confidentialité et l'anonymat. Les préoccupations suscitées par cette question comptent parmi les éléments principaux pris en compte par les répondants dans leur décision de répondre ou non à un questionnaire.

de participer, augmentant ainsi le taux de non-réponse. Toutefois, le risque en vaut la chandelle si l'objectif visé est de gagner la confiance de l'ensemble des participants éventuels à une enquête.

2.8 La transmission de l'information

En expliquant brièvement au répondant comment les informations qu'il fournit seront transmises, on lui permettra de mieux comprendre à quoi elles serviront. On pourra, à titre d'exemple, mentionner que les données ne seront présentées que sous forme de sommaires, ou qu'elles ne seront communiquées qu'à une organisation particulière, etc. Cet aspect pourrait être abordé en même temps que l'explication de la confidentialité des données puisqu'il existe un rapport logique entre ces deux composantes de l'enquête.

2.9 L'appréciation

Pour s'appuyer sur les autres composantes de la stratégie, les chercheurs ne devraient pas sous-estimer l'importance de démontrer aux répondants éventuels que leur participation à l'enquête est précieuse (Dillman, 1978). Cette appréciation du temps et des efforts consacrés par une personne est une mesure simple mais efficace d'encourager les gens à participer.

3. PREMIÈRE ÉTUDE DE CAS

3.1 Enquête nationale de 1995 sur les détenus

L'enquête nationale sur les détenus a été réalisée en automne 1995 par Price Waterhouse, pour le compte de Services correctionnels Canada (SCC). Price Waterhouse (PW) et son Centre national des sondages ont réalisé cette enquête auprès de plus de 4 000 détenus de 44 pénitenciers fédéraux du Canada. L'enquête a abordé une vaste gamme de sujets pour répondre à la fois aux exigences de la loi ainsi qu'aux recommandations d'un groupe de travail fédéral qui avait été chargé d'étudier les risques de santé (p. ex., le SIDA) auxquels sont exposés les détenus des pénitenciers fédéraux. Dix-neuf personnes spécialement formées ont été recrutées pour visiter ces établissements et faciliter la collecte des questionnaires remplis par les détenus et analysables à la machine.

Pendant les étapes de la préparation et de la conception du projet, PW a conçu une stratégie de communications complète. La tâche était particulièrement ardue compte tenu du nombre d'«intervenants» intéressés. Pour assurer la réussite de ce projet, PW avait besoin de faire connaître les objectifs et les mesures envisagées pour les atteindre (l'ensemble des composantes décrites dans les sections antérieures du présent document) non seulement aux répondants éventuels, les détenus, mais également aux comités de détenus, au personnel, aux gardiens et aux responsables régionaux que l'on désignera globalement sous le vocable de «gardiens». Le défi consistait à obtenir la participation de tous ces intervenants qui tous pouvaient devenir des sources possibles de non-réponses.

La mise au point de la stratégie de communications a suivi un cheminement itératif, PW fournissant les ébauches initiales et demandant ensuite la participation successive de l'équipe de projet de SCC. On a établi un point de contact à chaque établissement pour assurer la direction et le contrôle de la collecte des données et de la communication de l'information (p. ex., lettres) et favoriser ainsi la bonne marche de ce projet complexe.

SCC a transmis des avis officiels à tous les gardiens et à tous les membres du personnel régional pour les informer des aspects logistiques de l'opération. Des lettres ont également été envoyées par PW à tous les points de contact ainsi qu'à tous les comités de détenus des établissements visés. Ces lettres abordaient les sept composantes stratégiques dont nous avons fait état plus haut. Ces lettres avaient pour but, à la limite, d'obtenir la participation de ces personnes. La participation des contacts serait nécessaire pour assurer l'accès aux répondants éventuels. On a cherché à obtenir la collaboration des comités de détenus en

jugeant que cela contribuerait à accroître le taux de réponse des détenus. Cette démarche comportait cependant un risque : celui de se voir refuser l'appui d'un comité de détenus, ce qui aurait conduit à un type inhabituel (ou «non traditionnel») de non-réponse, celui d'un refus d'autorisation. Dans ce cas précis, la décision de ne pas répondre aurait été prise sans l'intervention du répondant éventuel (dans une telle situation, il aurait pu être intéressant de déterminer dans quelle mesure ces répondants pouvaient être jugés hors champ). Une prise de contact supplémentaire a été organisée après la réception de la lettre grâce à des rencontres personnelles avec les contacts des établissements et les comités de détenus pour étudier les questions et les préoccupations et mettre le point final aux aspects logistiques de la collecte des données. Compte tenu des circonstances de ce projet de recherche, le succès de l'opération dépendait de la mise en oeuvre soignée d'une stratégie de communications.

Dans l'ensemble, le taux de non-réponse pour cette enquête a été établi à 34,6 %. Comme prévu, ce taux a varié énormément entre les établissements et entre les régions. Il a oscillé entre 0 % à 69,0 % pour les établissements, et de 29,9 % à 39,9 % pour les régions.⁴ Les motifs précis des non-réponses variaient, mais la plupart de ces dernières étaient dues au départ des détenus par suite de transferts ou de libérations (en particulier, dans les établissements à sécurité minimale) ou à un refus pur et net de collaborer.

Nous soutenons que sans l'adoption d'une stratégie de communications, les taux de non-réponse auraient été beaucoup plus élevés. En outre, l'inclusion des huit composantes décrites plus haut a contribué à régler les problèmes jugés importants par les répondants au cours du test pilote de cette recherche.⁵

4. DEUXIÈME ÉTUDE DE CAS

4.1 Enquête de 1995 sur les personnes touchant des prestations d'invalidité du Régime de pension du Canada

L'enquête de 1995 sur les personnes touchant des prestations d'invalidité du Régime de pension du Canada a été réalisée par Statistique Canada (SC) pour le compte de Développement des ressources humaines Canada (DRHC). Elle avait pour principal objectif d'identifier et d'établir le profil des Canadiens recevant des prestations d'invalidité en vertu du Régime de pension du Canada (RPC). Un échantillon aléatoire stratifié a été tiré au niveau régional afin de permettre l'analyse des différences régionales d'un bout à l'autre du Canada. Les questions examinées comprenaient les raisons du versement des indemnités, les caractéristiques démographiques des personnes et des familles et, surtout, la situation financière et les revenus des bénéficiaires. Les résultats de l'enquête étaient également liés aux dossiers administratifs du RPC (DRHC) aux fins de la réalisation d'analyses ultérieures plus détaillées et d'expériences de modélisation.

On a élaboré une stratégie de communications détaillée, dans ce cas, pour des raisons semblables à celles invoquées dans la première étude de cas. Toutefois, il convenait en plus d'aborder des questions très différentes dans cette première prise de contact. On a présumé au départ que les répondants de cette étude hésiteraient à aborder la question de leurs prestations de pension, de peur de se les voir retirer. Il fallait donc tenir compte de cette préoccupation. Par ailleurs, comme nous avions l'intention de recueillir des informations détaillées sur les revenus et la situation financière des répondants, il nous fallait prendre le temps de préparer ces derniers à nous communiquer ces informations. Un point de contact (un numéro 800) a également été établi, et beaucoup des répondants l'ont utilisé. Il semble qu'on ait surtout voulu vérifier la validité de l'étude ou exprimer des craintes sur le sort réservé aux prestations de pension. Après avoir obtenu des renseignements supplémentaires sur l'étude et l'utilisation prévue des données, les répondants se sont montrés plus enclins à participer.

⁴ Services correctionnels Canada. (1995). Rapport de recherche. Gouvernement du Canada.

⁵ Services correctionnels Canada. (1994). Rapport d'étude pilote. Gouvernement du Canada.

Le taux global de non-réponse de cette enquête a été établi à 45,9%.⁶ Les taux de réponse ont été sensiblement inférieurs chez les répondants plus jeunes. Nous avons distingué trois niveaux de non-réponse, dont une partie correspondait au refus de remplir le questionnaire entier. Près de la moitié de l'ensemble des non-réponses ne pouvait, selon nous, être attribuée au «contact préalable» ni à la mise en oeuvre de la stratégie de communications. Sur l'ensemble des cas de non-réponse, 60,7 % n'avaient pas rempli le questionnaire ou ne l'avaient rempli que «partiellement», 23,9 % n'avaient pas de numéro de téléphone correspondant à l'adresse indiquée dans le dossier administratif même après les efforts déployés pour les retracer et 15,4 % n'avaient pas répondu à la question finale concernant leur consentement éclairé, réponse dont on avait besoin pour lier les données de l'enquête aux données administratives du RPC aux fins de l'analyse détaillée.

Malgré un taux de non-réponse que certains pourraient qualifier d'élevé comparativement à ce qu'on obtient habituellement à Statistique Canada, le recours à une stratégie de communications pour ce projet a été jugé utile. La question à l'étude était à la fois controversée et sensible, et l'examen de la ventilation des non-réponses a permis d'observer que près de la moitié d'entre elles pouvaient être attribuées à des raisons sur lesquelles la stratégie n'avait pas de prise.

5. INCIDENCES

5.1 Résultats d'une stratégie de communications

À ce point-ci de la discussion, on pourrait prétendre qu'un moyen efficace de déterminer l'incidence statistique véritable d'une stratégie de communications, telle que nous la définissons dans le présent document, consisterait à l'adopter comme partie intégrante du plan expérimental. À défaut de disposer de preuves concrètes ou empiriques à l'appui de cette hypothèse, nous devons examiner brièvement les avantages et les inconvénients perceptibles du recours à une telle stratégie dans le cadre de l'application d'une méthode de collecte de données.

5.2 Avantages

5.2.1 Établissement de rapports efficaces avec les répondants au début du processus de collecte

Le défi constant du responsable d'enquête consiste à gagner et à maintenir la confiance des répondants. À défaut d'une telle confiance et d'une compréhension des enjeux, les répondants seront moins enclins à participer. La réussite de tout projet de collecte de données dépend essentiellement d'une transmission des enjeux importants, de la détection précoce des problèmes et des préoccupations des répondants et du recours rapide aux solutions voulues.

5.2.2 Augmentation des taux de réponse

La production de données valides et rigoureuses compte parmi les principaux objectifs de toute enquête. Cela va de soi. Or, des taux de réponse élevés contribuent largement à la réalisation de cet objectif. Même s'ils ne permettent pas, en soi, de garantir une réduction du biais causé par la non-réponse, d'éviter le recours à des interprétations complexes et à la pondération pour expliquer une variable, et d'accroître sensiblement la confiance que les chercheurs et les analystes portent aux résultats obtenus, ils y contribuent certainement. La stratégie de communications n'est pas une fin en soi, mais c'est un outil qui aide les chercheurs à communiquer d'importants messages aux répondants. Ces messages, s'ils sont bien

⁶ Statistique Canada (1995). Enquête de 1995 sur les personnes touchant des indemnités du Régime de pension du Canada; rapport final. Gouvernement du Canada.

compris, favoriseront l'établissement d'un climat de confiance essentiel pour promouvoir la participation des répondants.

5.2.2 Raccourcissement de la période de collecte

Dans notre discipline comme dans beaucoup d'autres, «le temps c'est de l'argent». Un des objectifs du recours à une stratégie de communications consiste à répondre à certaines des préoccupations et des interrogations des répondants avant d'entamer l'étape de la collecte des données, pour ainsi permettre aux interviewers de gagner du temps. Nous ne prétendons pas que les stratégies de communications sont peu coûteuses ni qu'elles peuvent remplacer les rapports entre l'interviewer et le répondant. Toutefois, nous sommes d'avis qu'un tel contact préalable permettra de réduire le temps que les interviewers devront utiliser pour régler des questions liées aux objectifs de la recherche et à l'utilisation des données. Ainsi, le recours à une stratégie de communications risque selon nous d'influer sur les recettes réalisées.

5.3 Inconvénients

Toute démarche, qu'elle soit théorique ou pratique, comporte des limites. Ainsi, notre modèle ne s'applique pas à toutes les situations envisageables. Nous énumérons ci-après certains des inconvénients que peut poser une telle stratégie.

5.3.1 Risques d'un refus précoce de collaborer

L'établissement d'un contact préalable pourrait permettre aux participants de refuser plus facilement de collaborer. En contactant les répondants éventuels plus tôt et en leur fournissant le nom ou le numéro d'une personne-ressource, on risque de leur permettre de se désister plus facilement. Comme c'est le cas pour tous les refus possibles, il faut chercher à dissuader les non-participants éventuels. On peut comparer cette stratégie à celle qui consiste à ne pas demander aux répondants éventuels s'ils ont le temps de remplir un questionnaire.

5.3.2 L'augmentation de la non-réponse conduit à un risque de biais

On a déjà indiqué que le contact préalable peut conduire au refus immédiat de collaborer. Toutefois, l'incidence des non-réponses ne doit pas être négligée. Les non-réponses doivent être classées et analysées avant l'analyse des données, afin d'en mesurer l'incidence, ainsi qu'à l'étape de l'interprétation des résultats et de la préparation du rapport. À cette fin, dans tous les cas où on compte recourir à une stratégie de communications, il conviendra d'en examiner les incidences perceptibles ou prévisibles sur le taux de réponse.

6. AMÉLIORATION DE LA STRATÉGIE

6.1 Comment améliorer la stratégie de communications

Nous avons la faculté de pouvoir nous adapter et de faire l'expérience de diverses méthodes de rechange. Il en va de même pour la stratégie de communications. Si une composante ne donne pas les résultats attendus ou ne convient pas à un groupe particulier de répondants, nous avons tout loisir de la modifier. Se faisant, nous pouvons étudier des solutions de rechange afin, si possible, d'améliorer notre stratégie pour finalement accroître les taux de réponse et diminuer d'autant les taux de non-réponse. Quatre grandes questions méritent selon nous d'être examinées si nous souhaitons améliorer encore ce type de stratégie. Premièrement, les chercheurs doivent continuer à étudier les solutions de rechange pour promouvoir la participation des répondants et accroître le taux de réponse. Quelle autre solution pourrait-on envisager, hors des méthodes traditionnelles, pour accroître les taux de réponse et abaisser les taux de non-réponse? Deuxièmement, nous devons examiner l'exactitude et la véracité des informations

transmises aux répondants. Chaque projet exigera que l'on réponde à des besoins différents en cette matière et les questions à aborder lors du contact préliminaire varieront en conséquence. Il importe donc de revoir constamment les objectifs et les buts précis poursuivis. Troisièmement, il importe d'insister sur la confidentialité des informations. On peut définir cette notion de diverses façons, mais la clarté et l'exactitude sont essentielles si nous tenons à ce que les répondants saisissent bien le concept et acceptent de ce fait de participer. Finalement, nous devons examiner les moyens différents d'établir un tel contact préalable. Traditionnellement, nous avons eu recours à la poste. Toutefois, la multiplication des choix en matière de communications pourrait nous offrir d'autres moyens de contacter les répondants éventuels.

7. CONCLUSION

Le débat sur l'utilité de contacter les répondants avant l'administration d'un questionnaire se poursuivra. Une des façons de convaincre les chercheurs consiste évidemment à fournir des preuves empiriques de l'utilité d'une stratégie de communications (voir Dillman, 1978). À mesure que nous cherchons à accumuler un volume suffisant de telles preuves empiriques, nous devons toutefois continuer à chercher d'autres solutions. La concurrence dans la recherche sur les enquêtes et la recherche du répondant idéal est là pour rester. Il ne suffit plus de compter sur la réputation d'une entreprise ni de présumer que les gens prendront le temps de répondre à nos appels. Les chercheurs doivent se pencher sur des méthodes de rechange qui se démarqueront des méthodes traditionnelles. Avec les nouvelles technologies qui pointent à l'horizon et les développements prévisibles de l'Internet, la concurrence ne fera qu'augmenter. Si nous partageons nos connaissances, nous contribuerons à faire en sorte que les répondants partagent plus volontiers leurs informations.

8. BIBLIOGRAPHIE

Blalock, Hubert M. (1982) *Conceptualization and Measurement in the Social Sciences*. Beverly Hills: Sage Publications.

Buckland, William R. et Fox, Ronald A. (1963) *Bibliography of Basic Texts and Monographs on Statistical Methods, 1945-1960*. Édimbourg et Londres : Oliver and Boyd.

Church, Allan H. (1993) "Estimating the Effect of Incentives on Mail Survey Response Rates." *Public Opinion Quarterly*. Volume 57:62-79. American Association for Public Opinion Research.

Services correctionnels Canada (1994). Rapport de l'étude pilote. Gouvernement du Canada.

Services correctionnels Canada (1995). Rapport de recherche. Gouvernement du Canada

De Vaus, D.A. (1991) *Surveys in Social Research*. Third Edition. Sydney: Allen and Urwin.

Dillman, Don A. (1978) *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons.

Goyder, John. (1987) *The Silent Minority: Nonrespondents on Sample Surveys*. Cambridge: Polity Press. Cambridge.

Mills, Frederick C. (1955) *Statistical Methods*. New York: Holt, Rinehart and Winston.

Rubin, Donald B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Statistique Canada (1995). Enquête de 1995 sur les personnes touchant des prestations d'invalidité du Régime de pension du Canada; rapport final. Gouvernement du Canada.

FAVORISER LA RÉPONSE AUX ENQUÊTES SUR L'AGRICULTURE

Patricia Whitridge¹

RÉSUMÉ

Encourager les agriculteurs à répondre aux enquêtes de Statistique Canada n'a cessé de prendre de l'importance ces dernières années. En effet, il s'agit d'une population finie et très asymétrique, comme on en rencontre couramment dans les enquêtes sur les entreprises. Recueillir l'information pertinente des plus gros exploitants s'avère virtuellement impossible. Le fardeau imposé aux propriétaires d'une ferme de plus petite envergure est également considérable. Étant donné les sommes que le gouvernement investit dans la gestion de l'agriculture et l'intérêt concentré sur ce secteur, il était inévitable qu'on réalise de nombreuses enquêtes. Pour ces raisons, on s'est efforcé d'alléger le fardeau du répondant et d'encourager les agriculteurs à répondre aux sondages.

Parmi les mesures adoptées jusqu'à présent, on retrouve un recours accru aux données administratives sur l'impôt, la coordination des enquêtes présentant des périodes de référence similaires, la plus grande réduction possible du chevauchement des principales enquêtes et des tentatives délibérées pour améliorer la qualité de la base de sondage, afin de mettre un terme aux prises de contact superflues. On a aussi fait appel à des groupes de réflexion afin de mieux comprendre les préoccupations des agriculteurs à l'égard des enquêtes de Statistique Canada, recouru à des méthodes de collecte différentes, selon les préférences du répondant, et noué des liens personnels avec les propriétaires des plus grandes exploitations agricoles. Dans quelques cas, les répondants reçoivent des publications sommaires sur l'agriculture canadienne.

Le présent document décrit plus en détail les initiatives prises à ce titre et met en relief les enquêtes touchées, en plus de présenter quelques résultats.

MOTS CLÉS : Enquêtes agricoles; réponse aux enquêtes.

1. INTRODUCTION

Statistique Canada a pour mission de recueillir régulièrement des données agricoles auprès des agriculteurs canadiens afin de fournir des statistiques sur la superficie et le rendement des grandes cultures et des cultures de fruits et de légumes, sur le bétail et sur les finances. Toutes les exploitations agricoles du Canada sont visées, peu importe leur taille, leur emplacement ou le type de production. Environ 17 enquêtes agricoles ont lieu chaque année ou sur une base infra-annuelle. On entreprend aussi tous les ans des enquêtes spéciales de plus petite envergure en vue de recueillir des renseignements d'un intérêt particulier pour ceux qui les parrainent. En 1995, on avait établi environ 220 000 contacts au sein d'une population de 280 000 agriculteurs. Il convient néanmoins de préciser que beaucoup d'agriculteurs ont été approchés plus d'une fois. Le coût total de la collecte des données était de 2 900 000 \$ en 1995, ce qui n'est pas négligeable.

Dans la majorité des cas, les enquêtes se déroulent au téléphone selon la technique de l'interview téléphonique assistée par ordinateur (ITAO). Les tailles d'échantillon varient de plusieurs centaines à 30 000 répondants et les périodes de collecte durent d'une à plusieurs semaines. Le taux de réponse pour les enquêtes effectuées par ITAO varie de 89 % à 99 %.

La base de sondage de la plupart des enquêtes agricoles vient du recensement quinquennal de l'agriculture. On considère que la liste du recensement est fixe pour cette période de cinq ans; dans les

¹ Patricia Whitridge, Division des méthodes d'enquêtes-entreprises, Statistique Canada, Ottawa (Ontario) Canada K1A 0T6.

années intercensitaires, les principales enquêtes s'appuient sur une base aréolaire pour compenser le taux de natalité et les changements aux exploitations agricoles existantes. D'un point de vue pratique, les données du recensement sont chargées sur le Registre des fermes, banque de données administratives sur les agriculteurs.

Globalement, le programme d'enquêtes agricoles constitue un lourd fardeau pour les agriculteurs canadiens. En effet, la majorité des exploitations sont des fermes familiales qui comptent peu d'employés, voire aucun. Pour ces entreprises, c'est l'exploitant principal ou un membre de la famille qui doit prendre le temps de répondre aux questions. Le degré d'instruction des agriculteurs variant considérablement, la qualité des données recueillies en souffre, surtout lorsqu'il s'agit d'enquêtes financières.

On n'a pas réalisé de progrès énormes en vue d'améliorer le taux de réponse aux enquêtes agricoles de Statistique Canada. Ces dernières années, on a plutôt pris diverses initiatives de moindre envergure, chacune s'adressant à un groupe précis d'exploitations agricoles. Ce sont ces initiatives que nous examinerons ici.

2. RECENSEMENT DE L'AGRICULTURE DE 1996

Tous les cinq ans, Statistique Canada entreprend le Recensement de l'agriculture, le plus récent ayant eu lieu en mai 1996. À cette occasion, on recueille des renseignements sur tous les aspects de l'agriculture, notamment la superficie des cultures, le bétail, les finances élémentaires, les pratiques de gestion de l'entreprise et l'exploitation des terres. Le recensement porte sur toutes les fermes qui écoulent leurs produits sur le marché, peu importe l'importance du chiffre d'affaires. Les données du Recensement de l'agriculture servent de base de sondage aux enquêtes permanentes.

Le nombre de réponders téléphoniques augmente depuis quelques années dans les exploitations agricoles. Certains agriculteurs possèdent désormais plusieurs téléphones : un à la maison, un autre (parfois même un numéro différent) à l'étable ou au bureau et, à l'occasion, un téléphone cellulaire dans le tracteur. Ces changements ont permis de communiquer avec l'agriculteur quand il se trouve au champ, ce qui a accru le nombre de contacts. On recueille tous les renseignements sur les numéros de téléphone pour l'ensemble des agriculteurs responsables d'une exploitation donnée; lors des recensements précédents, on ne saisissait que l'information sur l'exploitant principal, qui était ensuite transmise aux intervieweurs. Ces renseignements complémentaires devraient faciliter la communication avec l'agriculteur ou un de ses associés.

Après transfert des données du Recensement de l'agriculture au Registre des fermes, on devrait noter une amélioration quant au nombre d'agriculteurs rejoints, en partie à cause des données complémentaires recueillies. D'autre part, si on parvient à le rejoindre à un moment où il n'est pas trop occupé, lorsqu'il conduit le tracteur par exemple, l'agriculteur pourrait être plus enclin à fournir les renseignements désirés. Auparavant, on communiquait souvent avec lui au moment du repas, lorsqu'il était à la maison, si bien qu'il pouvait se montrer plus réticent.

3. REMANIEMENT DES ENQUÊTES

Statistique Canada remanie les enquêtes agricoles en cours tous les cinq ans, au terme du Recensement de l'agriculture. L'actualisation concerne principalement les enquêtes sur les cultures et le bétail, qui ont lieu plusieurs fois par année. D'autres enquêtes de plus petite envergure, notamment celles sur les fruits et les légumes, la serres, les gazonnières et les pépinières, et la pomme de terre, subissent aussi des modifications après chaque recensement. On examine diverses facettes du plan d'échantillonnage, par exemple la base de sondage, les méthodes de stratification, la répartition et la sélection de l'échantillon et son maintien durant la période concernée (Denis et Whitridge, 1995). On teste les méthodes d'estimation afin de tirer parti du plan d'échantillonnage et on en vérifie la robustesse dans le temps. D'autre part, il est important de tenir compte du chevauchement des échantillons des enquêtes sur les cultures et le bétail et d'autres enquêtes, car peu de temps sépare bon nombre d'entre elles.

Actuellement, les contraintes budgétaires revêtent un intérêt particulier, car il est essentiel de réaliser des économies durant les cinq années que dure le plan d'échantillonnage. C'est pourquoi on a beaucoup réfléchi à la possibilité de combiner le plan d'échantillonnage des enquêtes sur les cultures et le bétail ou, du moins, la collecte des données sur le terrain. Les deux approches ont déjà été testées et chacune présente des avantages et des inconvénients. On craignait notamment qu'une enquête combinée couvrant tous les aspects de l'agriculture exige un questionnaire trop long, qui découragerait les répondants. D'un autre côté, on ne serait plus contraint de communiquer avec les agriculteurs se spécialisant dans cultures pour ne recueillir des renseignements que sur leur bétail lors de l'enquête sur le bétail. On a étudié comment contrôler le chevauchement des deux échantillons, soit en maximisant ceux-ci en vue de combiner les deux enquêtes, soit en le réduisant le plus possible afin d'entreprendre des enquêtes distinctes. Chaque option pose des difficultés, car les deux échantillons se renouvellent annuellement durant la période intercensitaire de cinq ans.

On a préféré recueillir les données des deux enquêtes séparément et réduire au maximum le chevauchement des échantillons. Si la chose est réalisable, on diminuera la taille de l'échantillon et allégera le contenu du questionnaire à certaines occasions, afin d'encourager la participation. Les intervieweurs sur le terrain sauront si on est entré en contact avec le répondant pour une autre enquête peu de temps auparavant afin de le ménager davantage.

4. GRANDES EXPLOITATIONS

Ces deux dernières années, on a déployé des efforts particuliers pour résoudre les problèmes que soulève la collecte de l'information des grandes exploitations, telles que définies par l'importance de leur chiffre d'affaires. Quelques-unes des exploitations agricoles les plus importantes du Canada appartiennent à de grandes sociétés, telles McCain et Nabisco. Il est très difficile de recueillir des renseignements sur leurs activités agricoles dans le cours normal du processus d'interview. En effet, les grandes sociétés ne sont pas toujours en mesure de divulguer les données sur leurs produits dans le format que requiert l'enquête ou ne désirent pas toujours le faire. Par ailleurs, la personne chargée de répondre à une enquête au sein de l'organisation pourrait varier, selon le contenu du questionnaire.

On a mis sur pied un groupe à qui on a confié la tâche de communiquer avec chacune des grandes sociétés et exploitations agricoles en vue d'en établir le profil. L'exercice a permis d'identifier la personne avec qui il est préférable de communiquer pour chacune des enquêtes et d'établir la meilleure façon de l'aborder (Blais, 1995). Ces renseignements sont utilisés lors de la collecte des données dans le cadre des enquêtes agricoles depuis à peu près un an maintenant, notamment pour le Recensement de l'agriculture.

Dans le cas des enquêtes permanentes, c'est toujours la même personne qui communique avec les grandes exploitations au téléphone. Cette personne se familiarise vite avec l'entreprise et la manière dont les données sont fournies. Un lien personnel prend naissance entre l'intervieweur et le répondant, ce qui

accroît la confiance et incite ce dernier à fournir l'information souhaitée. L'expérience acquise l'an dernier porte à croire qu'on obtient des renseignements de meilleure qualité et un taux de réponse supérieur, surtout quand on sait qu'auparavant, les entreprises de ce genre avaient peu tendance à répondre aux enquêtes courantes.

Différents scénarios ont été envisagés au moment d'élaborer un modèle applicable à la collecte de données d'entreprises aussi grandes que complexes. Ainsi, il a fallu choisir entre communiquer avec l'exploitation une fois par année au moyen d'un questionnaire omnibus et communiquer avec elle aussi souvent que nécessaire pour les enquêtes. La première option aurait exigé la conversion des données reçues en un format utilisable pour chaque enquête. Après consultation auprès des grandes exploitations, on a opté pour la deuxième solution.

Le taux de réponse des grandes entreprises s'est établi à environ 99 % lors du Recensement de l'agriculture de 1996, une amélioration considérable comparativement à ce qui se produisait dans le passé.

5. PROVINCES DE L'ATLANTIQUE

Jusqu'à présent, le programme de statistique agricole de Statistique Canada couvrait les dix provinces. Terre-Neuve fait l'objet d'enquêtes distinctes, principalement à cause du très petit nombre de fermes qu'on y trouve. La qualité des estimations obtenues pour les autres provinces de l'Atlantique (Île-du-Prince-Édouard, Nouveau-Brunswick et Nouvelle-Écosse) est toutefois inégale. Pour certaines enquêtes et certains produits, on obtient de très bons coefficients de variation (CV). Dans d'autres cas, ces derniers laissent à désirer. Étant donné le petit nombre d'exploitations agricoles dans ces provinces, le fardeau de réponse est très élevé. Une augmentation de la taille des échantillons en vue d'accroître la fiabilité des données ne constitue donc pas une solution valable.

Pour cette raison, on a décidé de modifier les enquêtes au moyen des données du dernier recensement de l'agriculture et d'établir une base de sondage en vertu de laquelle les provinces de l'Atlantique feraient l'objet d'un programme d'enquêtes distinct. Des consultations ont été amorcées avec des spécialistes des provinces de l'Atlantique en vue d'élaborer un programme d'enquêtes qui n'alourdira pas exagérément le fardeau de réponse tout en améliorant la qualité et l'utilité des données.

Le programme envisagé réduira le nombre d'enquêtes, car on se servira d'un plan d'échantillonnage combiné pour les cultures et le bétail en plus d'intégrer la collecte des données pour les enquêtes sur les producteurs d'œufs et la pomme de terre (Denis, 1996). On prélèvera un échantillon important en novembre afin de recueillir les chiffres définitifs sur les superficies ensemencées, la production et les inventaires de bétail, on obtiendra des données sur les produits importants au niveau infrarégional et on entreprendra une petite enquête de suivi en juillet. De cette façon, on devrait réduire considérablement le nombre de contacts et obtenir pour la première fois des données infrarégionales, comme on l'a demandé lors des consultations.

Le nouveau programme d'enquêtes pour les provinces de l'Atlantique sera introduit peu à peu, la première étape devant survenir à l'automne de 1997.

6. DONNÉES ADMINISTRATIVES

Depuis le début des années 80, les statistiques agricoles fournies par Statistique Canada comprennent des totalisations reposant sur les données administratives, dans le cadre du programme de données fiscales. L'information vient de Revenu Canada et porte sur un échantillon d'agriculteurs dont on transcrit les états financiers détaillés. Ces données permettent d'effectuer une analyse des revenus et des dépenses. Certaines fermes fournissent aussi un bilan, ce qui permet d'en estimer l'actif et le passif, bien que la qualité des

données varie, l'agriculteur n'étant pas tenu de fournir un bilan avec des états financiers. Les changements apportés aux formulaires qui servent à signaler les revenus et les dépenses d'une entreprise amélioreront l'opportunité et la qualité des données fiscales diffusées, car l'information sera disponible pour les particuliers qui rapportent le revenu d'un travail autonome et pour les sociétés. Statistique Canada a participé aux négociations avec Revenu Canada concernant la disposition exacte des formulaires. Les changements envisagés seront introduits au cours des deux prochaines années. En raison de la profusion des données fiscales, Statistique Canada n'aura plus besoin d'entreprendre une enquête pour recueillir des précisions sur les revenus et les dépenses.

Le projet de Données complètes sur les fermes a été lancé au début des années 90. Son but était de rassembler et d'intégrer les renseignements sur tous les aspects de l'agriculture - les variables des cultures et bétail, ainsi que les postes financiers (Wlútridge et Ménard, 1994). On s'est efforcé de rapprocher les données des enquêtes sur les produits et les données administratives de nature financière afin de brosser un tableau complet de la ferme. Diverses techniques statistiques ont été examinées dans le cadre de cet exercice, toujours dans l'espoir d'éviter une enquête distincte pour recueillir de l'information, d'une part, sur les variables des produits et, d'autre part, sur les finances. Les données du projet ont été offertes à différents intervenants qui exploitent l'information fournie sans qu'on note pour autant un alourdissement du fardeau des répondants.

Agriculture et Agroalimentaire Canada a besoin d'information pour appliquer efficacement divers programmes de soutien agricole dans le cadre de son mandat. Dans la mesure du possible, on recourt aux données recueillies par les organismes gouvernementaux en vertu des programmes d'assurance et des services d'inspection, par exemple. En l'absence de données administratives, il arrive qu'on procède à de petites enquêtes.

7. ENQUÊTE FINANCIÈRE SUR LES FERMES

L'Enquête financière sur les fermes est une enquête bisannuelle entreprise conjointement par Statistique Canada et Agriculture et Agroalimentaire Canada. Son but consiste essentiellement à recueillir des renseignements qui serviront à analyser la situation financière des agriculteurs canadiens. Les résultats permettent d'évaluer les programmes de subvention agricole et d'établir quels changements y apporter. On recueille des renseignements sur le bilan détaillé et quelques données sur les revenus et les dépenses. La méthode de collecte est complexe, puisqu'on commence par écrire aux agriculteurs échantillonnés pour leur présenter l'enquête et en expliquer l'objet. Des intervieweurs communiquent ensuite avec les répondants par téléphone pour savoir quel moment se prête le mieux à une plus longue interview en personne. Enfin, on répond au questionnaire qui repose largement sur les comptes agricoles.

Autrefois, le taux de non-réponse pouvait atteindre jusqu'à 20 % pour certaines variables importantes. On attribuait le taux de refus particulièrement élevé à la nature délicate des questions et à la longueur de l'interview. On s'est donc efforcé de motiver les répondants, avec un succès mitigé. Les mesures d'encouragement variaient de la remise de calendriers et de stylos à l'envoi de résultats partiels de l'enquête aux intéressés. Dans l'espoir d'améliorer la situation, Statistique Canada a effectué une étude au moyen de techniques qualitatives pour établir de quelle manière accroître les réponses en prévision de l'Enquête financière sur les fermes de 1993.

L'étude en question comprenait une analyse critique du questionnaire, des interviews personnelles, des interviews en profondeur et des groupes de réflexion composés d'agriculteurs des trois régions du Canada (Lawrence et Laffey, 1993). On a d'abord demandé à des spécialistes de la conception des questionnaires d'examiner le questionnaire de l'enquête et identifié les difficultés potentielles mises en relief lors des interviews et des discussions en groupes de réflexion. De la base de sondage, on a ensuite prélevé un échantillon d'agriculteurs qui ne feraient pas partie de l'échantillon de l'enquête de 1993. On a rejoint les participants et on les a conviés à répondre à une interview personnelle que sera suivie d'une discussion de

groupe quelques jours plus tard. Certains répondants qui n'ont pu assister à la discussion de groupe, ont subi une interview en profondeur.

Maintes recommandations sont ressorties de l'étude : modification de l'approche initiale à l'égard des répondants, révision de la manière dont les questions étaient formulées, adaptation du contenu du questionnaire et de l'enchaînement des questions, choix d'un meilleur moment pour la collecte de données et recours à des intervieweurs ne venant pas de la région, pour n'en citer que quelques-unes. Les résultats de l'étude se sont avérés utiles lors du développement de la méthode de collecte des données de l'Enquête financière sur les fermes de 1993. Les principales constatations pour Statistique Canada se rapportaient à la manière d'influencer la participation à l'enquête.

8. CONSULTATION DES UTILISATEURS

Statistique Canada recourt à maintes tribunes pour consulter les utilisateurs, surtout en ce qui concerne l'agriculture. On estime que ces consultations ont amélioré le taux de réponse, car plus de répondants participent au processus de l'enquête et prennent conscience de l'utilité des données.

Une rencontre annuelle entre les représentants fédéraux et provinciaux des organismes de statistique agricole sert de point de départ aux consultations. Il y a échange d'information sur les progrès récents dans le cadre des différents programmes et il est possible à cette occasion de poser des questions et de donner des commentaires sur le programme de statistique agricole de Statistique Canada.

Par ailleurs, des réunions ont été organisées avec de nombreux groupes d'intervenants, partout au pays, dans le cadre du projet de Données complètes sur les fermes. Ces réunions ont lieu à peu près tous les ans, parfois plus souvent, en fonction de l'avancement du projet. On demande aux participants quelles données les intéresseraient et quelles devraient être leurs caractéristiques sur le plan du contenu, du moment et des moyens de diffusion.

Enfin, il convient de mentionner le Conseil consultatif de la statistique agricole. Cet organisme se réunit chaque année pour discuter du travail de Statistique Canada à l'égard du programme de statistique agricole. Des spécialistes des quatre coins du pays, tant du monde universitaire que du secteur privé, donnent leurs commentaires et critiques, et décortiquent les programmes de Statistique Canada. On soulève parfois des questions auxquelles il faut répondre lors de réunions subséquentes.

9. CONCLUSION

Le fardeau des répondants ayant augmenté ces dernières années, on a déployé beaucoup d'efforts pour élaborer des enquêtes plus conviviales quant au mode de collecte, au questionnaire et aux méthodes d'enquête. Dans le secteur agricole, de nombreuses mesures secondaires affectent presque tous les aspects du programme statistique : enquêtes, recensement et programmes reposant sur les données administratives.

Dans le cadre du processus de remaniement des enquêtes, on a tenu compte de la meilleure façon dont les données devraient être recueillies, que ce soit au moyen d'un plan d'échantillonnage unifié, d'un système de collecte intégré ou de deux plans distincts avec chevauchement minimal des échantillons. Le programme s'est assoupli, ce qui a facilité l'adoption de procédures spéciales pour les grandes exploitations agricoles et l'implantation d'un programme d'enquête distinct, mieux adapté, pour les provinces de l'Atlantique.

Lorsqu'on reçoit des demandes d'information, on étudie désormais sérieusement la possibilité de recourir aux données administratives avant d'envisager une nouvelle enquête. Des procédures font couramment

appel à des essais pilotes et à des groupes de réflexion lors de l'élaboration du questionnaire et des méthodes de collecte destinées à la nouvelle enquête. De cette façon, on s'assure que cette dernière, même si elle augmente le fardeau imposé à une population restreinte, ne nuira pas au taux de réponse si difficile à maintenir. Par ailleurs, on recourt de plus en plus à d'autres moyens pour recueillir le point de vue des intéressés, qu'il s'agisse des répondants et des utilisateurs des données.

Comme nous avons tenté de l'illustrer, on n'a pas réalisé de progrès marquants en vue d'encourager la réponse aux enquêtes agricoles. Cependant, bon nombre de petites initiatives relatives à différents programmes d'enquête ont amélioré et préservé les taux de réponse. Le personnel des bureaux régionaux et de l'administration centrale chargé de la coordination doit être félicité pour le travail qu'il a accompli afin de maintenir un taux de réponse si élevé.

10. BIBLIOGRAPHIE

Blais, K. (1995). Multi-Agriculture Operations, Large Farms, Special Farms and Specified Farms (MULES) Project, Présenté aux réunions fédérales-provinciales sur les statistiques agricoles, Ottawa, Ont.

Denis, J. (1996). Atlantic Canada - Survey Redesign Methodology Proposal, Rapport technique statistique interne, Ottawa, Ont.

Denis, J. et Whitridge, P. (1995) 1996 Post-Censal Redesign of Agricultural Surveys, Présenté au Comité consultatif sur les méthodes statistique. Octobre 1995, Ottawa, Ont.

Lawrence, D. et Laffey, F. (1993). Qualitative Testing of the Farm Financial Survey Questionnaire, Proceedings of the International Conference on Establishment Surveys, Buffalo, New York.

Whitridge, P. et Ménard, M. (1994). Methodological Issues Involved in the Creation of a Farm Level Database, Proceedings of the Annual Research Conference, Washington, D.C.

MINIMISATION DE LA NON-RÉPONSE DANS UNE ENQUÊTE PAR PANEL

Heather Laurie, Rachel Smith et Lynne Scott¹

RÉSUMÉ

Nous présentons dans le présent article une évaluation de certaines des méthodes et de certains des systèmes d'enquête utilisés dans le cadre de la *British Household Panel Study* (BHPS). Nous décrivons les méthodes utilisées par la BHPS pour le traitement de la non-réponse à l'aide de systèmes de maintien des panels de méthodes de dépistage des répondants déménagés et de conversion des refus pendant le travail sur le terrain. L'analyse porte sur des données provenant des quatre premières vagues de la BHPS, réalisées de 1991 à 1994; elle vise à déterminer les tendances longitudinales des réponses et des motifs des refus. Nous examinons les raisons invoquées par les répondants pour refuser de participer à une interview ou pour se retirer définitivement du panel. Nous décrivons la méthode de conversion des refus ainsi que les résultats obtenus. Finalement, nous examinons l'effet de la stabilité de l'intervieweur sur le maintien de la coopération des membres de l'échantillon. Nos résultats laissent conclure que dans le contexte d'une enquête longitudinale par panel fondée sur un ensemble relativement complexe de procédures, il est impératif de minimiser la non-réponse et de maintenir des taux de réponse élevés pendant la durée de l'enquête.

MOTS CLÉS : Taux de réponse; méthodologie de l'enquête longitudinale; méthodes de travail sur le terrain.

1. INTRODUCTION

La réalisation d'une enquête longitudinale par panel présente un certain nombre de problèmes particuliers qui ont une incidence directe sur la qualité des données recueillies. Pour les études longitudinales comme les enquêtes par cohorte ou par panel, la minimisation de la non-réponse visant à contrer les effets éventuellement dommageables de l'attrition et à maintenir un échantillon viable est essentielle (Kazprzyk et coll., 1989). La non-réponse d'enquête est reconnue depuis longtemps comme un phénomène complexe aux multiples facettes (voir par exemple Sudman et Bradburn, 1977). Même si les enquêtes longitudinales par panel partagent beaucoup des difficultés auxquelles font face les enquêtes transversales dans la recherche d'un taux de réponse élevé, elles présentent, par leur nature même, des difficultés supplémentaires liées aux exigences du taux de réponse. Les panels présentent deux problèmes principaux particuliers à leur conception et qui peuvent causer leur attrition graduelle. Le premier est la mobilité géographique des membres de l'échantillon. Si les répondants déménagent et deviennent impossibles à retracer malgré tous les efforts déployés, ils deviendront autant de sources de données perdues. Par ailleurs, les répondants qui présentent la plus grande mobilité géographique ont tendance à se distinguer, sous d'autres aspects, des répondants sédentaires. C'est ainsi qu'on observe un problème d'attrition différentielle où une catégorie particulière de répondants peut devenir sous-représentée à l'intérieur de l'échantillon. La seconde source de perte de données, qui est aussi plus généralisée, est due aux refus motivés très souvent par ce que nous appellerons la «fatigue du répondant». À chaque vague d'interviews, les répondants ont le choix de refuser de participer à l'enquête. Après avoir coopéré pour des périodes qui peuvent atteindre quelques années, les membres d'un panel peuvent devenir ennuyés, se désintéresser de l'étude ou simplement avoir l'impression qu'ils en ont «fait assez». Même si la plupart des répondants ont tendance à valoriser leur participation et à y prendre plaisir, il arrivera inévitablement que certains décideront d'interrompre leur collaboration. Comme les personnes mobiles, celles qui

¹ Heather Laurie, Rachel Smith, Lynne Scott, ESRC Research Centre on Micro-social Change, University of Essex, Colchester, CO4 3SQ, Royaume-Uni.

refusent d'être interviewées ont tendance à présenter des caractéristiques particulières qui risquent de produire des profils d'attrition différents et, dans le pire des scénarios, un biais à l'intérieur des données recueillies.² La présente étude a donc pour objectif d'évaluer l'efficacité des méthodes utilisées dans le cadre de la BHPS pour minimiser la non-réponse en fonction du temps.

2. LA BRITISH HOUSEHOLD PANEL STUDY

La *British Household Panel Study* est une enquête nationale par panel sur les ménages englobant plus de 10 000 personnes réparties entre quelque 5 500 ménages de Grande-Bretagne. Elle est réalisée par le *ESRC Research Centre on Micro-Social Change* de la *University of Essex*. L'échantillon englobe des résidences privées d'Angleterre, du pays de Galles et d'Écosse. La BHPS a été entreprise en septembre 1991 et les membres du panel sont interviewés tous les ans. Lors de la première vague de l'enquête, 13 840 personnes, comprenant les enfants âgés de moins de 16 ans, ont été dénombrées dans 5 511 ménages. De ce groupe, 9 912 adultes éligibles (c.-à-d., âgés de 16 ans ou plus) ont été interviewés et 352 interviews par personnes interposées ont été réalisées pour donner un taux de réponse supérieur (entrevues complètes avec au moins un membre du ménage) à 74 %. Les entrevues de la sixième vague de l'enquête ont commencé en septembre 1996, et les répondants recevront la visite des intervieweurs pour la septième fois en septembre 1997. Les informations recueillies par la BHPS portent à la fois sur les ménages et sur leurs membres. L'intervieweur se renseigne d'abord sur la composition du ménage, la propriété et le coût du logement, les produits de consommation et les dépenses du ménage consacrées au carburant et aux aliments. Il recueille ensuite une vaste gamme d'informations sur la migration, l'état de santé des membres et leur recours aux services de santé, des détails sur leur emploi et leur revenu, leurs idées reçues et leurs opinions, ainsi que sur l'organisation du ménage. Finalement, un questionnaire à remplir soi-même permet de recueillir des informations sur les attitudes et d'autres sur l'état de santé général des répondants (voir Rose et coll., 1991 pour une description complète du contenu et de la conception de la BHPS). L'interview sur le ménage dure une dizaine de minutes et chaque interview personnelle dure en moyenne 40 minutes; une interview complète pour un répondant donné demande donc moins d'une heure. En outre, depuis 1994, les enfants âgés de 11 à 16 ans et qui vivent dans les ménages de l'échantillon doivent remplir eux-mêmes un court questionnaire (Scott et coll., 1994). Même si l'objectif consiste à réaliser une interview complète avec chaque adulte éligible, on recueille également des informations par personnes interposées et on procède à de courtes entrevues téléphoniques afin d'obtenir des informations de base sur le plus grand nombre possible de membres de l'échantillon. Comme on l'a observé dans le cas d'autres panels, le recours à des outils de collecte des données souples et à un mélange de méthodes permet de maintenir le contact avec des membres de l'échantillon qui risqueraient autrement de se désister à la longue (Schupp et Wagner, 1996). Au cours des cinq premières années de la BHPS, les répondants qui se sont soumis à une entrevue complète ont reçu un bon-cadeau d'une valeur de 5 £ en guise de remerciement. Les jeunes participants ont pour leur part reçu un bon-cadeau de 3 £. Ces bons étaient postés aux répondants avec une lettre de remerciement et une carte d'avis de changement d'adresse après l'interview³. Par ailleurs, les intervieweurs distribuent également de petits présents comme des stylos marqués du symbole de l'enquête ou des petits agendas. Même si certaines observations portent à conclure que ces encouragements augmentent le taux de réponse dans les catégories marginales, en particulier chez les personnes à faible revenu comme les personnes âgées vivant seules, on utilise ce

² La question de l'attrition différentielle dans l'échantillon et celle des méthodes de pondération utilisées par la BHPS pour tenir compte du biais de non-réponse ne sont pas directement examinées dans le présent article. Prière de consulter A. Taylor (1993) pour obtenir une description des caractéristiques des répondants en rapport avec l'attrition différentielle et les méthodes de pondération.

³ À partir de la vague 6 de l'enquête, on a commencé à poster le bon-cadeau à l'avance aux répondants qui avaient collaboré l'année précédente et à ceux ayant franchi le cap des 16 ans et étant de ce fait devenu éligibles à l'interview des adultes. Les intervieweurs ont des bons-cadeaux qu'ils peuvent remettre directement à tous les autres répondants au moment de l'entrevue. La valeur du bon-cadeau a également été portée de 5 à 7 £ par adulte interviewé et de 3 à 4 £ par jeune interviewé.

moyen principalement pour remercier les répondants de leur collaboration et non en guise de rémunération pour le temps qu'ils nous ont consacré.

Toutes les études par panel adoptent les règles suivantes qui permettent de désigner quels sont les membres de l'échantillon à inclure dans le suivi et dans quelles circonstances chaque personne devrait ou non être incluse (Burgess, 1989; Kalton et Lepkowski, 1985). Les règles de suivi et d'éligibilité qui sont adoptées dans le cadre du plan d'enquête global imposent certaines contraintes sur la façon dont le suivi est assuré d'une année sur l'autre et nécessitent la mise en place d'un système de gestion de l'échantillon relativement complexe. Dans le cas de la BHPS, les membres de l'échantillon sont suivis même s'ils quittent leur ménage d'origine, créent un nouveau ménage ou rejoignent un ménage dont ils faisaient partie lors d'une vague antérieure. Tous les membres échantillonnés à la vague 1 de l'enquête sont désignés membres permanents de l'échantillon (PSM) et sont suivis lorsqu'ils déménagent, même s'il s'agit d'enfants âgés de moins de 16 ans. Lorsque les enfants ont 16 ans, ils deviennent éligibles aux interviews. Les nouveaux membres du ménage sont inclus dans l'échantillon et sont éligibles à l'interview lorsqu'ils continuent à partager ce ménage avec un membre permanent de l'échantillon. L'échantillon de la BHPS est un échantillon de personnes; ce sont les personnes qui sont suivies dans leurs déplacements d'un ménage à l'autre. Comme Duncan et Hill (1985) l'ont fait remarquer, il n'existe pas de ménages longitudinaux; il n'y a que des personnes longitudinales. Les ménages changent leur composition, de nouveaux ménages sont formés et d'autres disparaissent sous l'effet combiné des déplacements des personnes; la personne est donc la seule unité utile pour la description d'un contexte longitudinal.

3. TAUX DE RÉPONSE LONGITUDINAUX

Dans une enquête par panel, la question de la méthode à utiliser pour faire la description longitudinale des taux de réponse peut devenir quelque peu problématique. Pour chaque vague transversale de l'enquête, on peut calculer le taux de réponse du ménage pour l'ensemble des ménages sélectionnés ou pour l'ensemble des ménages contactés si on inclut les nouveaux ménages créés pendant la durée des interviews. On peut de la même manière calculer les taux de réponse individuels pour chaque vague transversale. Même si les taux de réponse transversaux peuvent donner une idée du succès ou de l'échec de chaque période d'interviews, ils nous renseignent peu sur les profils longitudinaux de la réponse, sur l'ensemble des années que dure l'enquête. Il n'est pas possible non plus d'évaluer l'incidence de l'attrition graduelle. Ainsi, pour obtenir cette perspective longitudinale, nous avons besoin de calculer les taux de réponse au niveau individuel, d'une vague sur l'autre. Nous travaillons en effet avec un échantillon de personnes qui se déplacent d'un ménage à l'autre, ce qui complique le calcul des taux de réponse des ménages d'une vague à l'autre. Les répondants pour lesquels on possède un dossier d'interviews complet, c'est-à-dire ceux qui ont subi une interview complète à chacune des vagues de l'enquête, sont ceux qui fournissent les informations longitudinales de base nécessaires à la réalisation d'analyses par panel significatives. Une des façons d'évaluer les taux de réponse d'une vague à l'autre consiste à examiner les taux de réinterviews d'une vague à l'autre au niveau individuel (voir tableau 1). Sur les 9 912 répondants qui ont subi une interview complète à la vague 1 de l'enquête, 87,7 % de ceux qui étaient toujours éligibles ont été réinterviewés à la vague 2. À la vague 3, 90,4 % des répondants éligibles de la vague 1 et interviewés à nouveau à la vague 2 ont été réinterviewés. À la vague 4, 94,8 % des répondants éligibles ont été réinterviewés. Même si ces taux de réponse aux réinterviews peuvent sembler élevés comparativement à beaucoup de taux de réponse transversaux, ils ont été jugés nécessaires pour maintenir un échantillon longitudinal viable. Sur les 9 912 répondants de la première vague de l'enquête, 7 131 possèdent un dossier d'interviews complet pour la période d'étude de 4 ans. Ceci signifie qu'après l'exclusion des personnes devenues inéligibles en cours de route, nous avons retenu 74 % de notre échantillon original sur lequel nous possédons des informations en continu pour les quatre premières années de l'enquête.

Tableau 1 : Taux de réponse aux réinterviews individuelles de la BHPS — 1991 - 1994-

	1991	1992	1993	1994
Répondants originaux	9 912			
Répondants de la V 1 réinterviewés		8 568 (87,7 %)	7 629 (90,4 %)	7 131 (94,8 %)
Répondants de la vague précédente réinterviewés		(S.O.)	8 216 (90,2%)	8 278 (94,0 %)
Nombre total d'interviews		9 459	9 032	9 062

* Interviews individuelles complètes des répondants éligibles à chaque vague.

Outre l'échantillon longitudinal de base, nous possédons dans beaucoup de cas des informations sur des répondants interviewés à une ou plusieurs occasions au cours de l'enquête, mais pas à toutes les vagues, et dont les données sont donc discontinues. Par exemple, 207 des répondants de la vague 1 ont subi une interview complète aux vagues 1, 2 et 4, mais non à la vague 3. De même, 173 répondants de la vague 1 ont subi une interview complète à toutes les vagues subséquentes, sauf à la vague 2. Selon le type d'analyses qu'on compte réaliser, ces répondants possèdent des informations longitudinales qui peuvent certainement nous être utiles. Certains des répondants à données discontinues appartiennent à des ménages échantillonnés à la vague 1 qui n'ont pas été interviewés à la vague 1 ou qui avaient moins de 16 ans à cette époque, et qui sont depuis devenus éligibles à une interview complète. À chaque vague de l'enquête, près de 200 jeunes ayant atteint 16 ans deviennent éligibles à une interview complète et nous cherchons à les interviewer tous. Au cours des trois années qui ont suivi la vague 2 de l'enquête, 432 des jeunes membres de notre échantillon original ont été interviewés. Les autres répondants pour lesquels nous possédons des informations discontinues sont des membres temporaires qui se sont joints à l'enquête depuis la vague 1. En dépit du fait qu'ils ne font pas partie de l'échantillon original, beaucoup d'entre eux ont participé à l'enquête pendant 2 ou 3 ans, fournissant ainsi des informations contextuelles précieuses pour n'importe quelle analyse longitudinale tout en augmentant la taille de l'échantillon d'ensemble soumis à l'analyse transversale. Lorsque nous nous penchons sur l'échantillon original des personnes interviewées en fonction du nombre de vagues où elles ont été interviewées et du type de données recueillies, nous constatons que le pourcentage des membres de l'échantillon qui ont donné une forme quelconque d'informations longitudinales au cours des quatre années de l'enquête est plus élevé que celui fondé sur les interviews complètes de l'ensemble des quatre vagues. Sur les 9 391 répondants interviewés de l'échantillon original qui étaient éligibles à l'interview de la vague 4, 80 % (7 573) ont été effectivement réinterviewés à cette vague. Cette proportion est portée à 83 % (7 805) lorsqu'on y ajoute les informations recueillies par personnes interposées ou au moyen d'interviews téléphoniques. Même si les informations recueillies par ces deux derniers moyens sont quelque peu limitées, elles nous permettent de conserver les répondants visés dans l'échantillon longitudinal des personnes interviewées.

4. MÉTHODES DE MINIMISATION DE LA NON-RÉPONSE

Un certain nombre de méthodes ont été adoptées dans le cadre de la BHPS pour minimiser autant que possible la non-réponse. Ces méthodes sont intégrées au processus d'enquête et certaines d'entre elles, comme le dépistage des répondants, se poursuivent entre les vagues. La gestion d'un panel nécessite la mise en œuvre d'une gamme de mesures de contrôle de la qualité qui visent toutes à maximiser la réponse et à recueillir des données de haute qualité. Pour la BHPS, on a recours uniquement aux services d'intervieweurs expérimentés qui ont déjà travaillé à des enquêtes à échantillon aléatoire. En outre, d'une vague de l'enquête à l'autre, on cherche autant que possible à attribuer à chaque intervieweur le même groupe de ménages. Tous les intervieweurs nouvellement recrutés participent à une séance d'informations de deux jours avant d'entreprendre leur travail tandis que ceux qui ont déjà participé à des vagues antérieures bénéficient d'un breffage d'une journée. Le travail sur le terrain est contrôlé de près du début à la fin, et les intervieweurs sont contactés une fois par semaine pour faire rapport de l'état actuel des ménages et de chacun de leurs membres. Les intervieweurs sont tenus de faire au moins six tentatives de

contact à chaque adresse, à des moments différents de la journée, avant de déclarer un non-contact. En outre, les ménages qui ont fait l'objet de six tentatives de contact ou plus au cours de la vague précédente sont connus de l'intervieweur. Le contenu, la conception et la longueur des questionnaires constituent également des facteurs critiques qui influent sur la collaboration des membres de l'échantillon. Toutefois, au-delà de ces éléments qui s'appliquent à n'importe quelle opération de collecte de données, la conduite d'une enquête par panel pose trois exigences particulières qui exigent qu'on lui consacre des ressources supplémentaires : i) maintien du panel; ii) procédures de dépistage des répondants déménagés; iii) mise en œuvre d'un programme de conversion des refus.

4.1 Maintien du panel

À la lumière d'une étude d'enquêtes par panel réalisées précédemment, on a décidé très tôt d'élaborer une base de données (*Panel Maintenance Database* ou PMDB) pour garder le contact avec les membres de notre panel. Il est prioritaire, pour toute enquête par panel, de maintenir des registres exacts et à jour des allées et venues de chaque membre de l'échantillon. On a décidé de faire de la PMDB une liste distincte de la liste des noms et adresses des membres de l'échantillon pour deux raisons. Il fallait premièrement tenir compte des exigences de confidentialité, non seulement pour se conformer au *UK Data Protection Act*, mais également pour respecter nos propres normes d'éthique et protéger nos répondants. Les répondants de la BHPS reçoivent l'assurance que leurs noms et adresses ne seront jamais liés aux informations qu'ils fournissent. En maintenant une liste de noms et d'adresses séparée de la base de données de l'enquête, on s'assure que les liens directs entre les deux ne pourront être faits que par un nombre limité de personnes autorisées. Deuxièmement, la PMDB est remise à jour entre chaque vague d'interviews et cette opération est plus facile à réaliser avec une base de données séparée de celle des données de l'enquête recueillies à chaque vague. En élaborant nos méthodes de maintien du panel, nous avons surtout cherché à garder le contact avec nos répondants par des moyens autres que l'interview elle-même. Pour ce faire, nous avons eu recours à un certain nombre de techniques, y compris :

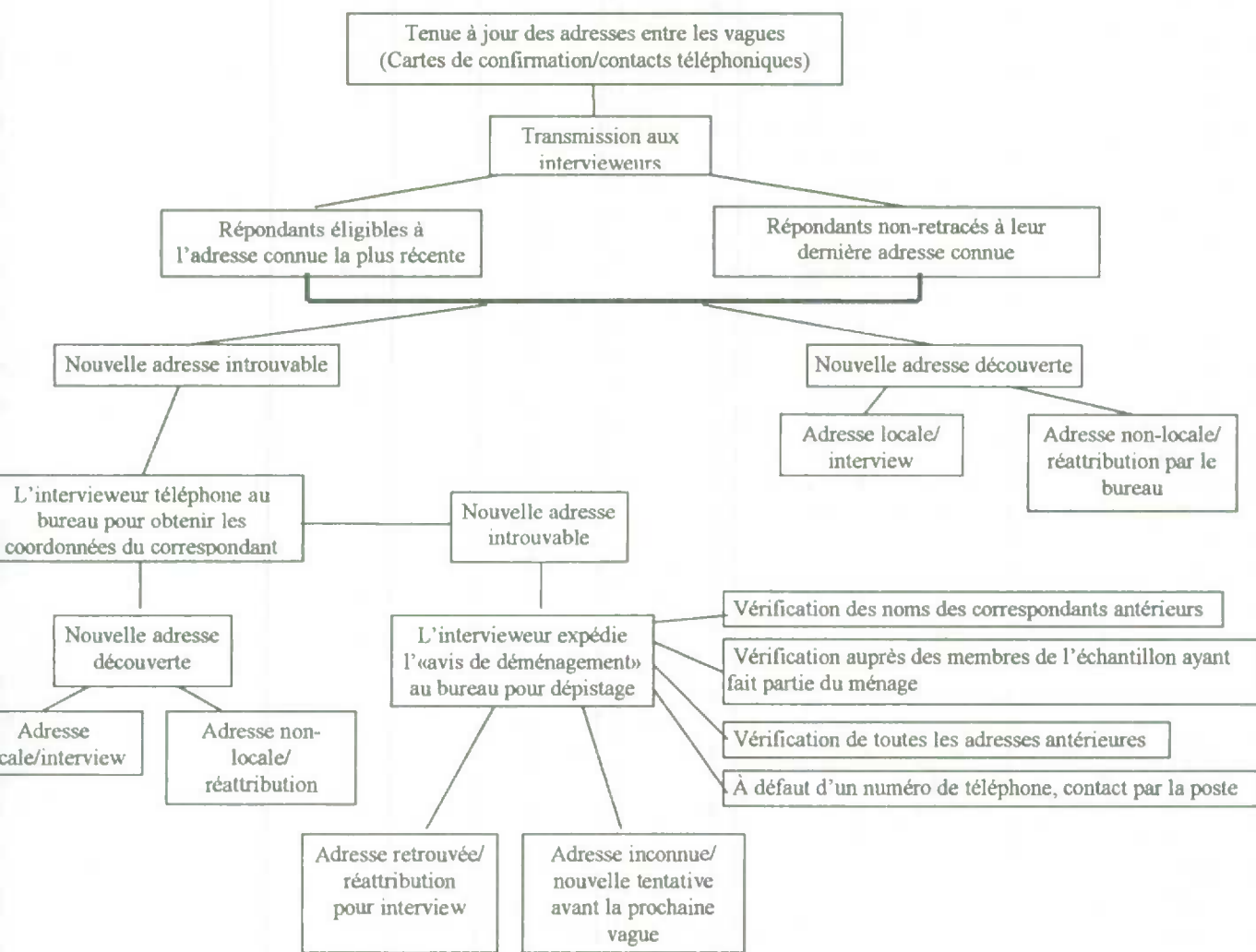
- désignation d'une personne-ressource, avec numéro téléphonique sans frais et service de répondeur, pour le bénéfice des répondants;
- prise en note des détails des contacts établis avec les répondants entre les vagues d'interviews;
- transmission à l'intervieweur des informations pertinentes concernant les répondants avant chaque vague d'interviews (p. ex., décès ou maladie d'un membre du ménage);
- diffusion, à l'intention des répondants, d'un court rapport annuel des résultats de la recherche et des activités en cours, accompagné d'une carte pré-affranchie de confirmation d'adresse;
- inclusion d'une carte d'avis de changement d'adresse avec les bons-cadeaux et les lettres de remerciement expédiées après chaque vague d'interviews;
- envoi d'un bon-cadeau de 5 £ à toute personne retournant une carte de changement d'adresse entre deux vagues d'interviews;
- tenue à jour de la liste d'adresses entre les vagues d'interviews;
- maintien d'une liste de toutes les adresses occupées successivement par chaque membre de l'échantillon;
- poursuite des opérations de dépistage des répondants pendant et entre les vagues d'interviews.

Nous sommes partis du principe que les répondants qui acceptent de poursuivre leur collaboration dans le cadre de notre enquête prennent un engagement qu'il convient de souligner, ils méritent d'être informés de l'usage que l'on fait des données recueillies. Des répondants nous ont indiqué qu'ils apprécient beaucoup de recevoir notre rapport annuel, que ce rapport leur donne le sentiment de contribuer à un projet utile et de jouer un rôle important dans la réalisation de l'enquête. En fait, beaucoup de répondants nous ont demandé de leur fournir des informations supplémentaires, demandes qui ont été traitées sur une base individuelle. Le maintien des contacts grâce à une correspondance postale entre les vagues de l'enquête procure aux répondants un sentiment d'appartenance et nous donne d'autre part l'occasion de tenir à jour nos listes d'adresses. Nous avons ainsi le loisir de vérifier l'exactitude de nos listes d'adresses non pas uniquement pendant les interviews mais également au cours des mois qui s'écoulent entre chaque vague, ce qui augmente encore l'efficacité des opérations de dépistage. Environ 500 cartes d'avis de changement d'adresse nous sont retournées chaque année et le tiers environ des répondants nous retournent leur carte de confirmation d'adresse avant le début de chaque vague annuelle d'interviews.

4.2 Dépistage des répondants déménagés

En procédant à la mise à jour des adresses entre les vagues d'interviews, on maintient une proportion maximale de ménages dont l'adresse est exacte au début de la vague. Burgess (1989) appelle ce procédé «*Forward tracing*». On a également recours à des méthodes rétrospectives, pendant les périodes d'interviews, lorsque l'intervieweur appelle, découvre qu'une personne a déménagé et qu'il cherche à découvrir sa nouvelle adresse. Chaque année, environ 10 % des membres de l'échantillon (1 000 personnes) déménagent. Dans près de la moitié de ces cas, nous recevons un avis quelconque du changement d'adresse soit grâce à la carte d'avis de changement d'adresse, à la carte de confirmation ou à un contact téléphonique. Pour le reste des répondants, la recherche commence au moment où l'intervieweur fait son premier appel et se trouve incapable de retracer le répondant. Les enquêtes par panel présentent l'avantage particulier de donner aux intervieweurs l'occasion de mieux connaître leurs répondants, de perfectionner leurs méthodes de dépistage et de se familiariser avec leur milieu d'étude, ce qui accroît leurs chances de retracer leurs répondants sans avoir à recourir à l'aide du bureau. Pour les cas inévitables où leurs recherches restent infructueuses, ils doivent remplir un «avis de déménagement» décrivant en détail les répondants dont on a perdu la trace. À chaque vague de l'enquête, les intervieweurs retournent au bureau de 200 à 250 de ces avis. Un des principaux moyens utilisés pour retracer ces personnes consiste à s'adresser aux personnes dont les noms ont été fournis par le répondant au cours des années précédentes. En effet, chaque année, on demande aux répondants d'indiquer le nom d'un correspondant qui risque d'être tenu au courant de leurs allées et venues. Cette méthode s'est avérée la plus efficace et la plus rentable pour retracer les répondants qui déménagent. Le répondant qui consent à nous fournir le nom d'une telle personne nous indique par ailleurs qu'il sera prêt à coopérer et à maintenir sa participation au cours des vagues futures d'interviews. De plus, au cours des années, nous parvenons à accumuler les noms de plusieurs correspondants, ce qui accroît d'autant nos chances de retracer les répondants qui déménagent. Le processus de dépistage est long et exige du personnel une certaine dépense de temps et d'énergie. On évalue à environ 10 £ par ménage le temps du personnel et les autres ressources comme les frais téléphoniques consacrés au dépistage des répondants déménagés. Toutefois, nous retraçons 50 % des ménages dont les intervieweurs ont perdu la trace, ce qui correspond à environ 125 ménages par vague. Dans le cas d'un panel longitudinal, ce nombre relativement restreint de ménages représente près de 2,5 % des ménages de l'échantillon, ce qui rend l'opération critique pour le maintien des taux de contacts et la réduction des pertes cumulatives avec le temps. Lorsque le coût de cette opération est réparti sur l'ensemble des 5 500 ménages de l'échantillon, le coût supplémentaire par ménage est inférieur 50 pences, soit une dépense qui en vaut largement la peine compte tenu du coût d'ensemble relativement élevé de l'enquête.

Figure 1 : Méthodes de dépistage de la BHPS



4.3 Conversion des refus

La minimisation du taux de refus est l'autre moyen important de maintenir des taux de réponse élevés. Même si on attache une importance primordiale à l'apprentissage, par les intervieweurs, de la façon d'aborder les répondants, il arrivera inévitablement que certains d'entre-eux refusent de participer. La mise en place d'un programme de conversion des refus est donc un élément important de réduction des pertes possibles dues aux refus. L'expérience a démontré que beaucoup des refus sont particuliers à la vague, c'est-à-dire qu'un répondant peut refuser de participer à cause de circonstances de nature immédiate qui pourraient ne plus se reproduire au cours des années suivantes. La deuxième catégorie de refus est celle des répondants qui décident de se retirer pour de bon de l'enquête. La question la plus difficile à résoudre dans le cadre d'une enquête par panel consiste à déterminer le point à partir duquel la décision d'un répondant de se désister est irrévocable, c'est-à-dire le moment à partir duquel l'éthique professionnelle nous interdit de tenter une conversion. Dans beaucoup de cas, nous avons observé qu'il est possible de faciliter le déroulement ou l'organisation de l'interview afin d'encourager le répondant à revenir sur sa décision et à participer à nouveau. Il s'agira dans certains cas simplement de parler au répondant, de lui fournir des explications plus détaillées sur les buts de l'enquête et de lui rappeler l'importance et le caractère irremplaçable de sa contribution. On peut en outre chercher à répondre aux besoins particuliers du répondant, lorsque cela est possible. Par exemple, le répondant pourra demander qu'on l'appelle à une date ou à une heure du jour particulières et on cherchera dans la mesure du possible à l'accommoder. Depuis la deuxième vague de l'enquête, on a mis sur pied un programme de conversion des refus qui fait maintenant partie intégrante de nos méthodes de travail. Chaque cas de refus signalé par un intervieweur est évalué comme suit :

- évaluation des motifs du refus;
- examen des antécédents du dossier et des contacts antérieurs avec le répondant;
- décision de tenter ou non une conversion;
- en cas de tentative de conversion, prise de contact par un intervieweur expérimenté;
- en cas d'accord, réattribution du dossier à un intervieweur;
 - réattribution à un intervieweur principal ou au surveillant de zone; à moins que le répondant ne demande à voir un intervieweur particulier;
 - nouvelle prise de contact par l'intervieweur dans les sept jours qui suivent la réattribution du dossier;
 - versement d'une prime à l'intervieweur pour toutes les interviews réussies par suite d'une conversion;
- en cas de refus du répondant de recevoir un intervieweur, réalisation d'une courte entrevue téléphonique;
- en cas de refus d'une entrevue téléphonique, nouvelle évaluation du dossier avant l'attribution à la vague suivante.

L'interview téléphonique est un mécanisme qui a été mis en place à la vague 3 pour conserver des répondants qui auraient autrement été irrémédiablement exclus de l'échantillon. Cette méthode a porté fruits puisque 50 % des 252 répondants qui ont accepté de participer à une interview téléphonique à la vague 3 ont accepté, à la vague 4, d'accorder une interview individuelle normale complète. On a recours au mécanisme de conversion des refus pour environ 300 ménages par vague. Dans presque 60 % des cas, on réussit à obtenir une interview individuelle complète ou une interview téléphonique. Comme dans le cas des méthodes de dépistage, ce nombre relativement petit de ménages récupérés représente entre 2 et 3 % des ménages de l'échantillon, soit une proportion sensible du total, vue sous l'angle des taux de réponses longitudinaux.

Pour traiter les cas de refus de la façon la plus adéquate, nous avons constaté qu'il était nécessaire de recueillir le plus d'informations possibles sur les motifs du refus et de tenir des dossiers des rapports entretenus avec les répondants ayant montré certaines réticences par le passé ou ayant présenté d'autres

problèmes lors d'une vague antérieure de l'enquête. À toutes les vagues, on a demandé aux intervieweurs de noter les raisons invoquées par les répondants pour refuser de participer au moment de l'interview. À partir de la vague 3, ces données ont été classées au bureau. À la vague 3, on a enregistré 719 refus de répondants qui «n'étaient plus intéressés» (24 %, n=173). Ces répondants appartenaient à des ménages où au moins une interview avait été réalisée par le passé et beaucoup d'entre eux avaient probablement le sentiment d'«avoir fait leur part». D'autres répondants ont attribué leur refus au fait qu'ils étaient occupés ou rarement à la maison (15 %, n=108) ou qu'ils étaient trop malades ou trop âgés (15 %, n=108). Dans 20 % des cas (n=144) aucun motif n'a été donné. À la vague 4, le nombre de refus a chuté du tiers environ, comparativement à la vague 3, pour atteindre un total de 475 ménages, une réduction partiellement due aux décisions prises entre les deux vagues sur la façon de traiter les refus. Lorsqu'un répondant refuse catégoriquement de prendre part à l'enquête, son nom est retiré de l'échantillon. Ainsi, chaque année, une certaine proportion de personnes résolument inintéressées à participer est retirée de l'échantillon, ce qui tend à accroître avec le temps la proportion des personnes disposées à collaborer. À la vague 4, la proportion des personnes qui ont refusé de participer parce qu'elles «n'étaient plus intéressées» était de 45 % (n=214), soit presque le double de la proportion enregistrée à la vague 3, même si en nombres absolus, le nombre des refus enregistrés à la vague 4 n'était que légèrement supérieur à celui enregistré à la vague 3. À la vague 4, les intervieweurs étaient devenus plus habiles à découvrir les raisons des refus et le nombre des refus inexplicables a été réduit des deux tiers ou presque. Au cours des séances de formation, on insiste auprès des intervieweurs sur l'importance de cette information pour l'amélioration des méthodes de conversion. Par ailleurs, la proportion des refus attribuées à des raisons personnelles ou familiales (p. ex., trop occupé, rarement à la maison, trop malade ou trop âgé) est demeurée la même à la vague 3 et à la vague 4.

Par ailleurs, il peut être intéressant d'examiner si certains refus peuvent être plus facilement convertis en interviews que d'autres. Nous présentons au tableau 2 les pourcentages des ménages de la vague 4 qui ont été convertis en fonction de la raison originale du refus et du revenu final du ménage après la tentative de conversion. À la vague 4, 276 ménages ont fait l'objet d'une tentative de conversion. Pour 26 % de ces ménages, au moins une interview individuelle complète a été réalisée. Une portion supplémentaire de 31 % des ménages ont accordé au moins une interview téléphonique tandis que 43 % ont refusé à nouveau. On risque davantage, dans l'ensemble, d'obtenir une interview téléphonique qu'une interview complète chez les ménages faisant l'objet d'un effort de conversion. Toutefois, ce choix dépend du type de motifs invoqués pour justifier le refus. La probabilité d'obtenir une interview téléphonique est plus élevée lorsque la raison du refus est liée au déroulement de l'enquête plutôt qu'à une raison personnelle ou familiale. Dans 69 % des cas où on a consenti à une interview téléphonique, le répondant avait soulevé des objections concernant le processus de l'enquête lui-même ou avait déclaré ne plus être intéressé. Par contre, la probabilité d'obtenir une entrevue complète est plus élevée si la raison originale du refus est liée à des problèmes personnels ou familiaux. Parmi les répondants qui ont finalement accepté d'accorder une interview, 37 % avaient attribué leur refus original à des raisons personnelles, comparativement à 22 % des répondants qui ont finalement accordé une interview téléphonique. En outre, les répondants dont le refus était motivé par des problèmes liés à l'enquête plutôt que par des problèmes personnels seront davantage portés à refuser une seconde fois. Dans le cas des refus liés à l'enquête elle-même, il importe au plus haut point de prendre les mesures voulues pour éviter, par exemple, des problèmes d'horaires ou de confidentialité. Pour éviter que les membres du panel ne se désintéressent de l'enquête, il convient également de leur rappeler l'importance des questions étudiées et l'incidence qu'elles peuvent avoir sur leur propre vie, afin d'entretenir leur intérêt. Il importe en outre d'apporter un soin particulier à la prise en compte des besoins et de la situation particulière des répondants lorsqu'on prend contact avec eux afin d'éviter les désistements dus à un simple manque de souplesse dans les méthodes utilisées et l'organisation du travail sur le terrain.

Tableau 2 : Issue de la conversion en fonction du type de motif du refus — vague 4, BHPS

Motif premier refus	Issue de la conversion		
	Int. complète %	Int. tél. %	Second refus %
Motifs liés à l'enquête	43,7 (31)	69,4 (59)	51,7 (62)
Motifs personnels ou familiaux	36,6 (26)	22,4 (19)	38,3 (46)
Autres/Refus non motivé	19,7 (14)	8,2 (7)	10,0 (12)
Total	25,7 (71)	30,8 (85)	43,5 (120)

Seuil de sign. : < 0,01

5. STABILITÉ DES INTERVIEWEURS

La stabilité des intervieweurs est un des principes que nous nous efforçons d'appliquer dans l'organisation du travail d'enquête de la BHPS. Des indices de nature anecdotique nous portent à croire que les répondants autant que leurs intervieweurs apprécient cette stabilité. Les répondants sont ainsi en mesure d'apprendre à connaître leur intervieweur, et de bâtir avec lui des liens de confiance mutuelle. L'intervieweur est pour sa part mieux en mesure de maintenir le contact avec des gens et des familles auxquels il voue un intérêt réel. Il est possible d'examiner l'effet de la stabilité de l'intervieweur sur la volonté des répondants à coopérer tout au long de l'enquête par panel. Depuis le début de l'enquête, 97 % des répondants ont conservé le même intervieweur pendant au moins 2 des 4 premières vagues. Parmi les répondants qui ont conservé la même adresse pendant toute la durée du panel, 46 % ont conservé le même intervieweur pendant les quatre vagues et 18 % ont conservé le même pour trois des quatre vagues. Nous indiquons au tableau 3 la proportion du nombre total des répondants de la vague 1 qui ont accordé une interview individuelle complète à chacune des vagues, selon qu'ils ont ou non déménagé entre deux vagues et selon qu'ils ont ou non conservé le même intervieweur. Un déménagement signifiera dans beaucoup de cas un changement d'intervieweur à la vague suivante, en particulier lorsque le répondant déménage hors de sa localité d'origine. Ce changement d'intervieweur est un des facteurs qui pourraient inciter les gens à se désister. Toutefois, les données du tableau indiquent que l'effet négatif le plus évident d'un changement d'intervieweur s'observe parmi la population des répondants sédentaires. Les taux de réponse des répondants sédentaires qui conservent le même intervieweur vague après vague sont toujours plus élevés que lorsqu'on a changé l'intervieweur. Pour les répondants qui ont déménagé, le changement d'intervieweur n'a pas cet effet. Il est possible que les répondants qui déménagent s'attendent d'une certaine façon à devoir changer d'intervieweur. Par contre, les répondants sédentaires s'attendent peut-être pour leur part à pouvoir garder le même intervieweur année après année.

Tableau 3 : Proportions des répondants de la vague 1 ayant accordé une interview individuelle complète selon qu'ils ont ou non déménagé et qu'ils ont ou non conservé le même intervieweur — BHPS, 1992 - 1994 (tous les répondants éligibles localisés)

		Choix de l'intervieweur par rapport à la vague précédente		
		Le même %	Différent %	Total %
Vague 1/2	Sédentaire	97,0 (4 784)	96,5 (3 040)	96,8 (7 824)
	Déménagé	96,6 (313)	98,2 (431)	97,5 (744)
Vague 2/3	Sédentaire	92,7 (5 360)	90,6 (1 737)	92,2 (7 097)
	Déménagé	95,1 (386)	93,9 (355)	94,5 (741)
Vague 3/4	Sédentaire	95,9 (5 616)	84,9 (1 184)	93,8 (6 800)
	Déménagé	96,6 (453)	97,6 (324)	97,0 (777)

Tableau 4 : Régression logistique de la probabilité pour un répondant de la vague 1 d'appartenir à un ménage non-répondant à la vague 4 (la catégorie entre parenthèses est la catégorie de référence)

	B	Seuil de sign.	Probabilité d'appartenir à un ménage non-répondant à la vague 4
Variables de coopération			
Coopération partielle du ménage à la V 1 (Coopération complète à la V 1)	0,3177	0,0000**	1,3739
Plus d'un intervieweur (Même intervieweur à toutes les vagues)	0,4630	0,0000**	1,5889
Observations de l'intervieweur			
Répondant passable à mauvais à la V 10,2234 (Répondant bon à très bon à la V 1)		0,0013*	1,2504
Pas de correspondant à la V 1 (Nom d'un correspondant à la V 1)	0,1750	0,0022*	1,1913
Problèmes de santé ou de langue à la V 1 (Pas de problème à la V 1)	0,1620	0,0266	1,1759
Conditions démographiques et socio-économiques			
Propriétaire du logement (Locataire du logement)	-0,4306	0,0000**	0,6501
Revenu mensuel du ménage	-9,0E-05	0,0217	0,9999
* Seuil de sign. < 0,01			
** Seuil de sign. < 0,001			

n=7 123 répondants de la vague 1 éligibles à toutes les vagues, sans information manquante/population sédentaire

Variables non significatives incorporées dans le modèle : présence ou absence d'enfants dépendants dans le ménage; degré de scolarité le plus élevé, situation par rapport au marché du travail à la vague 1 et absence ou présence d'un membre du ménage dans l'expérience à échantillons superposés de la vague 2.

Outre l'effet de la stabilité de l'intervieweur observé d'une vague à l'autre, nous pouvons également calculer la probabilité pour un répondant de la vague 1 de faire partie d'un ménage non-répondant à la vague 4, en fonction d'un certain nombre de variables clés, y compris la stabilité de l'intervieweur au cours de la durée de vie du panel. Dans le modèle décrit au tableau 4, nous avons inclus certaines caractéristiques démographiques et socio-économiques de base des répondants ainsi que des informations transmises par l'intervieweur sur les réactions du répondant pendant l'interview. On demande aux intervieweurs de noter, après chaque interview individuelle, le degré de coopération du répondant et l'existence de problèmes de santé qui pourraient influencer sur l'interview, et d'indiquer si le répondant était disposé à fournir le nom d'un correspondant qui serait tenu au courant de ses allées et venues au cas où il déménagerait entre deux vagues.

On peut observer dans ce modèle que le recours à plus d'un intervieweur au cours des quatre années de l'enquête est un prédicteur significatif du passage du répondant dans un ménage non-répondant à la vague 4. Les répondants qui changent d'intervieweur au cours de la durée de l'enquête par panel ont 58 % plus de chances que ceux qui conservent le même intervieweur de se trouver dans un ménage non-répondant à la vague 4, résultat qui permet de confirmer l'utilité de conserver le même intervieweur aux fins du maintien des taux de réponse longitudinaux. Les observations de l'intervieweur à la vague 1 nous aident également à prévoir la non-réponse à la vague 4. Les répondants qui sont classés «passables à mauvais» à la vague 1, ceux qui refusent de fournir le nom d'un correspondant aux fins d'un éventuel dépistage et ceux qui éprouvent des problèmes de santé ou d'autres problèmes nuisant à l'entrevue ont tous une probabilité plus élevée de se trouver dans un ménage non-répondant à la vague 4. Ces résultats confirment l'importance de transmettre les informations recueillies sur les répondants aux intervieweurs

de la vague suivante de manière à ce qu'ils puissent adapter leur comportement et accroître leurs chances de maintenir le contact. En outre, la coopération dont fera preuve le répondant dépendra nettement de l'attitude des autres membres du ménage face à l'enquête. Dans les cas où un des membres du ménage a refusé de collaborer ou n'a pu être contacté à la vague 1, les répondants risquaient davantage d'appartenir à un ménage non-répondant à la vague 4. En ce qui a trait à la formation des intervieweurs, il convient d'insister sur l'importance de chercher à obtenir la collaboration de tous les membres du ménage. Ces efforts peuvent avoir une incidence sur les résultats de la conversion des refus.

6. CONCLUSION

Le maintien de taux de réponse élevés dans le contexte d'une enquête longitudinale par panel nécessite la mise en place d'un ensemble passablement complexe de procédures et de systèmes d'enquêtes dont quelques-uns seulement ont été abordés dans le présent article. Certaines des procédures sont liées au travail sur le terrain et mises en application directement par les intervieweurs, tandis que d'autres, comme le maintien du panel, les méthodes de dépistage et de conversion des refus utilisées dans le cadre de la BHPS relèvent plutôt du bureau. L'expérience de la BHPS permet jusqu'à maintenant de démontrer clairement que les efforts supplémentaires consacrés au maintien du contact avec les membres du panel, au processus de conversion des refus et à la mise en place de méthodes de travail sur le terrain axées sur la satisfaction des besoins du panel sont justifiés. La combinaison de toutes ces méthodes a eu une incidence globale significative sur la minimisation de l'attrition et sur le maintien des taux de réponse à un niveau qui a permis d'assurer la viabilité de l'échantillon et la collecte de données de haute qualité nécessaires à une analyse de fond.

Nous remercions le Economic and Social Research Council (Royaume-Uni) et la University of Essex pour leur appui. Le travail décrit dans le présent article a été réalisé dans le cadre du programme scientifique du ESRC Research Centre on Micro-Social Change de Grande-Bretagne.

Nous remercions également Graham Kalton et Nick Buck pour les commentaires utiles qu'ils ont formulés sur les ébauches du présent article.

7. BIBLIOGRAPHIE

- Sudman, S. et Bradburn, N.M. (1977). *Response Effects in Surveys*. Aldine Publishing Co., Chicago
- Burgess, R.D. (1989). «Major issues and implications of tracing survey respondents» in Kasprzyk et coll. (éds), *Panel Surveys*. Wiley, New York.
- Corti, L. et Campanelli, C. (1992). «The Utility of Feeding Forward Earlier Wave data for Panel Studies» in Westlake et coll., *Survey and Statistical Computing*. North Holland, London.
- Duncan, G.J. et Hill, M.S. (1985). «Conceptions of Longitudinal Households. Fertile or Futile?» *Journal of Economic and Social Measurement* 13:361-375.
- Kalton, G. et Lepowski, J. (1985). «Following rules in SIPP». *Journal of Economic and Social Measurement* 13:319-329.
- Kasprzyk, D., Duncan, G., Kalton, G. et Singh, M.P. (éds) (1989). *Panel Surveys*. Wiley, New York.

Rose, D. et coll. (1991). «Micro-Social Change in Britain». *Working Paper No 1*, ESRC Research Centre on Micro-Social Change, University of Essex.

Schupp, J.P. et Wagner, G.G. (1996). «Maintenance of Long-On-Going Panel Studies — The case of the German Socio-Economic Panel Study (GSOEP)». Article présenté à la 4^e Conférence internationale sur la méthodologie des sciences sociales, Essex '96, University of Essex, Juillet 1996. Texte photocopié.

Scott, J., Brynin, M. et Smith, R. (1994). «Interviewing Children in the British Household Panel Survey». Article présenté au 5^e Symposium du *Research Committee on Empirical Family Research*, Amsterdam, Décembre 1994 (texte photocopié).

Taylor, A. (1993). «Sample characteristics, attrition and weighting» in Buck, N. et coll. (éds), *Changing Households : The BHPS 1990 to 1992*. ESRC Research Centre on Micro-Social Change, University of Essex.

SESSION 4

CONCEPTION DE QUESTIONNAIRE ET CONTRÔLE DE LA QUALITÉ

RÉDUCTION DES ERREURS NON DUES À L'ÉCHANTILLONNAGE PAR LA RECHERCHE COGNITIVE DANS LE SUPPLÉMENT DE LA *CURRENT POPULATION SURVEY* SUR LA RACE ET L'ETHNICITÉ

Ruth B. McKay¹

RÉSUMÉ

La recherche cognitive a joué un grand rôle dans le supplément de la *Current Population Survey* sur la race et l'ethnicité de mai 1995. Les interviews de recherche qui ont servi à tester les versions successives du questionnaire ont permis de réduire le taux d'erreur de l'enquête par l'identification et la rectification des problèmes soulevés par le questionnaire. Par ailleurs, le fait de contrôler les enquêtes durant la semaine de collecte des données de mai 1995 et l'analyse des réponses aux questions ouvertes sur le groupe multiracial et l'ascendance ou l'origine ethnique ont permis d'interpréter quelques constatations ambiguës résultant de l'analyse statistique des données du supplément. Selon la définition du terme « multiracial » utilisée dans la recherche, on est ainsi parvenu à distinguer les répondants « multiraciaux véritables » et « multiraciaux indéterminés (faussement positifs) ». Le groupe le plus important de répondants multiraciaux indéterminés se compose de personnes qui signalent deux ethnicités ou plus, p. ex. Écossais et Italien. Les résultats de la recherche cognitive révèlent qu'il sera difficile d'éliminer les erreurs non dues à l'échantillonnage dans les enquêtes sur la race et l'ethnicité tant que les concepteurs et les répondants ne s'entendront pas sur le sens donné aux expressions comme « race » et « ethnicité/origine ethnique ».

MOTS CLÉS: Erreurs de mesure; recherche cognitive; termes raciaux.

1. INTRODUCTION

Lessler et Sirken divisent les erreurs non dues à l'échantillonnage en plusieurs groupes : erreurs associées à la base de sondage, erreurs attribuables aux données manquantes sur certains éléments de l'échantillon et erreurs de mesure, ou « d'observation » comme on dit parfois, relevées lorsque les mesures obtenues ne rendent pas la valeur véritable des variables représentées. Parmi les erreurs de mesure, on trouve celles résultant de facteurs associés au questionnaire, à la manière dont les données sont recueillies, à l'intervieweur et au répondant (Lessler et Sirken, 1985).

La recherche cognitive entreprise par le *Bureau of the Census* consécutivement aux recensements décennaux de 1980 et de 1990 avait fait ressortir l'origine des erreurs de mesure dans les deux enquêtes (Martin et coll., 1990; Elias-Olivares et Farr, 1991; KISSAM et coll., 1993). Dans les années 90, le *Bureau of Labor Statistics* et le *Bureau of the Census* ont commencé à recourir à la recherche cognitive dans l'espoir de réduire les erreurs non dues à l'échantillonnage, notamment les erreurs de mesure, dans les enquêtes démographiques (McKay et de la Puente, 1996; Bates et coll., 1994).

Contexte

Les normes fédérales sur la classification raciale et ethnique établies en 1977 par l'*Office of Management and Budget* (OMB) établissent les catégories qui suivent : (race) blanc, noir, amérindien ou autochtone de l'Alaska; asiatique ou natif du Pacifique; (ethnicité) origine hispanique et origine non hispanique. Depuis l'adoption de ces normes, les citoyens qui fournissent des renseignements sur eux-mêmes et les utilisateurs des statistiques fédérales ont signalé que les catégories en question avaient perdu de leur utilité

¹Ruth B. McKay, Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Washington, DC 20212, É.-U.

face à la plus grande diversité de la population. En 1994, l'OMB a organisé un atelier de deux jours à la *National Academy of Sciences* afin de vérifier l'utilité des normes raciales et ethniques.

Selon les chiffres du recensement des États-Unis, le nombre d'enfants multiraciaux est passé de 460 302 à 1 937 496 entre 1970 et 1990 aux États-Unis (Bennett, McKenney et Harrison, 1995). D'autre part, des recherches du *Bureau of the Census* indiquent qu'en apportant deux changements aux catégories de race déclarées, on pourrait diminuer le sous-dénombrement des personnes hispaniques lors du recensement décennal. En ajoutant la catégorie « hispanique » et en renversant l'ordre dans lequel on s'enquiert de la race, puis de l'ethnicité du répondant, on réduirait sensiblement le nombre de personnes d'origine hispanique qui se classent dans la catégorie « autre » quand on leur pose la question initiale sur la race, si bien qu'on pourrait sauter la question sur l'ascendance hispanique au recensement décennal (Bates, et coll., 1993). Le multiculturalisme grandissant de la population soulève aussi de nouvelles préoccupations au sujet de l'interprétation des expressions: race, ethnicité, ascendance et origine nationale, surtout chez les personnes d'origine hispanique (Elias-Olivares et Farr, 1991; Kissam, Herrera et Nakamoto, 1993).

L'atelier de l'OMB de 1994 a débouché sur la constitution d'un comité consultatif interagence représentant plus de 30 organismes fédéraux, à qui on a confié la tâche de revoir les normes ethniques et raciales. Un groupe de travail a été mis sur pied pour évaluer de nouvelles catégories raciales et ethniques, ainsi que pour examiner les effets éventuels de tels changements sur les statistiques fédérales. La population a participé à l'exercice par le biais d'audiences publiques qui se sont déroulées dans les principales régions des États-Unis en juillet 1994. Le premier projet de recherche a été soumis vers la fin du même mois. Il s'agissait d'un supplément à la *Current Population Survey* (CPS) devant permettre la collecte de renseignements sur plusieurs grandes questions à l'étude.

LE SUPPLÉMENT DE LA CPS SUR LA RACE ET L'ETHNICITÉ

La CPS, une enquête mensuelle (du *Bureau of Labor Statistics* et du *Bureau of the Census*) sur 60 000 ménages représentant la population civile non institutionnalisée des États-Unis, recueille couramment des renseignements sur la race et l'ascendance des membres du ménage, le premier mois de l'interview. En testant de nouvelles questions sur la race et l'ethnicité sur la même population, on pourrait comparer la façon dont les répondants s'identifient à un groupe selon le libellé actuel et selon sa version remaniée. Ainsi, l'addition d'une catégorie « multiraciale » permettrait de mesurer le déplacement éventuel par rapport aux groupes existants, p. ex., noir, amérindien vers « multiracial ».

Parmi les sujets de recherche envisagés dans le cadre du supplément de la CPS, on retrouve les suivants : (1) incidence de l'inclusion d'une catégorie multiraciale à la liste des groupes raciaux; (2) effet de l'addition du terme « hispanique » à la liste de races; (3) incidence de la question sur l'origine hispanique quand elle précède la question sur la race; (4) questions devant établir l'acceptabilité d'autres termes, p. ex. afro-américain pour noir, latin pour hispanique; (5) questions d'ordre conceptuel destinées à nous en apprendre davantage sur les critères dont les répondants se servent pour se répertorier et classer les autres dans tel ou tel groupe racial, ethnique ou ancestral.

Le questionnaire du supplément a été rédigé par le groupe de travail à partir des commentaires de spécialistes en conception de questionnaires et d'experts des domaines spécialisés. Le second groupe se composait d'universitaires considérés comme des autorités sur les principaux groupes raciaux et ethniques des États-Unis.

Le supplément a été structuré en quatre panels ou versions :

Panel I: Questions distinctes pour la race et l'origine hispanique; pas de catégorie multiraciale.

Panel II: Questions distinctes pour la race et l'origine hispanique; catégorie multiraciale.

Panel III: Question combinée pour la race et l'origine hispanique; pas de catégorie multiraciale.

Panel IV: Question combinée pour la race et l'origine hispanique; catégorie multiraciale.

En dehors des différences relatives aux questions initiales sur la race et l'origine hispanique, les questions sur les autres aspects liés à la race et à l'origine ethnique, comme l'ascendance et les termes à connotation raciale préférés, seraient identiques pour les quatre panels. La dernière partie du supplément comprenait aussi les mêmes questions de nature conceptuelle.

2. MÉTHODOLOGIE

La recherche a été effectuée par une équipe multiraciale et multi-ethnique tirée de divers organismes gouvernementaux et universités (1). Le plan expérimental prévoyait le couplage de la race et de l'origine hispanique entre le répondant et le chercheur. Ce plan a pu être respecté pour tous les groupes sauf celui des amérindiens, qui ont été interrogés par un chercheur américain d'origine asiatique. L'équipe était également multidisciplinaire puisqu'on y retrouvait des anthropologues, des psychologues et des sociologues. La recherche s'est poursuivie de novembre 1994 à avril 1996.

2.1. Élaboration du questionnaire

Le protocole de recherche destiné à permettre l'évaluation du supplément prévoyait une interview personnelle, de vive voix, durant laquelle l'intéressé répondrait à toutes les questions des quatre panels à l'essai. Cela fait, le répondant devait paraphraser chaque question, c'est-à-dire expliquer en ses propres termes à l'intervieweur ce qu'elle signifiait. Dans le cas des questions intégrant des termes d'un intérêt particulier pour la recherche (p. ex. race, ethnicité, latin), le répondant était aussi prié de définir les termes utilisés.

Le matériel dont on s'est servi pour la recherche cognitive, y compris les quatre panels du supplément à tester et les protocoles des interviews de recherche, a été traduit en espagnol de manière à pouvoir être utilisé avec les répondants hispaniques. Pour la traduction, on a recouru aux services de traducteurs mexicains, portoricains et cubains. Les divergences entre les trois traductions ont été réconciliées lors d'une conférence à laquelle ont participé les trois traducteurs.

Le plan expérimental supposait le testage du supplément auprès des membres des groupes suivants : blanc; noir; amérindien et autochtone de l'Alaska; hispanique; asiatique ou natif des îles de Pacifique; multiracial. Les répondants ont été recrutés par le truchement des centres communautaires desservant les groupes ethniques et raciaux sur lesquels portait la recherche. Les interviews cognitives se sont déroulées en trois phases.

Phase un

Vers le milieu de novembre 1994, on a entrepris vingt interviews cognitives avec des répondants de la région métropolitaine de Washington. Le nombre d'interviews a permis de mettre les quatre panels du supplément à l'essai pour tous les groupes raciaux et ethniques sauf celui des Amérindiens et des Autochtones de l'Alaska, qui faisaient partie de la phase deux. Les interviews de la phase un ont fait ressortir maintes difficultés au niveau des quatre panels, notamment: 1) des questions vagues ou imprécises; 2) des questions de nature délicate; 3) des questions abstraites; 4) des problèmes de vocabulaire; 5) des effets d'ordre et 6) des redondances. Nous examinerons quelques-unes de ces difficultés dans ce qui suit.

Questions vagues ou imprécises

Q.4a : Vous avez choisi [race] dans la liste que je vous ai lue. Existe-t-il un autre groupe, plus précis, auquel vous vous associez?

Dans l'affirmative

Q. 4b Comment s'appelle ce groupe?

La question ne précisait pas qu'on s'intéressait à un groupe *ethnique* en particulier. Les réponses ont donc varié de « chrétien », à « adolescent récalcitrant », en passant par « franc-maçon » et « musulman ».

Q. 8a : Plus tôt, je vous ai lu une liste de groupes comprenant les catégories suivantes : blanc; noir; Amérindien, Esquimau ou Aléoute; asiatique ou natif des îles du Pacifique; et autre. Auriez-vous préféré trouver la catégorie « multiracial » dans cette liste?

Les répondants ont eu tendance à interpréter cette question comme une question générale et à donner une réponse « politiquement correcte » du genre « bien sûr, il faudrait l'ajouter à la liste ». Les répondants n'ont pas compris qu'on leur demandait s'ils auraient aimé signaler leur appartenance au groupe « multiracial ».

Questions abstraites

Q. 5 : D'après vous, quelle est la caractéristique la plus importante qui définit une race?

Q. 6 : Existe-t-il selon vous une distinction entre les termes race, ethnicité et ascendance?

La majorité des répondants ont jugé les questions conceptuelles trop ardues, y compris ceux qui avaient poursuivi en partie des études collégiales. Les répondants estiment que les questions sont trop abstraites et étaient persuadés qu'elles avaient pour but de mettre leur intelligence à l'épreuve. Plusieurs les ont trouvées particulièrement difficiles, car elles ne comportaient aucun « indice » ni liste d'éléments parmi lesquels choisir une réponse.

Problèmes de vocabulaire Très peu de répondants connaissaient le sens du mot « ethnicité ». Plusieurs ont cru que la question portait sur l'éthique de diviser la population en races. Une femme était persuadée que le terme « caractéristique » avait trait à son caractère.

Effets d'ordre Dans les panels un et deux, on demandait au répondant s'il aurait préféré avoir la catégorie « multiracial », après l'avoir interrogé sur son ascendance ou son origine ethnique.

Q. 5 : Quelle est votre ascendance ou origine ethnique?

Q. 8 a : Plus tôt, je vous ai lu une liste comprenant les catégories suivantes : blanc; noir; Amérindien, Esquimau ou Aléoute; asiatique ou natif des îles du Pacifique; et autre. Auriez-vous préféré trouver la catégorie « multiracial » dans cette liste?

Les répondants qui avaient signalé un aïeul de race différente à la question sur l'ascendance ont demandé s'ils devaient déclarer avoir une origine multiraciale, étant donné leur réponse à la question sur l'ascendance. Cette réaction n'a pas été notée avec les panels 3 et 4 où la question sur l'ascendance *suivait* celle sur la catégorie « multiracial ».

Redondance/Sensibilité Les expressions du genre « race », « ethnicité », « ascendance » et « origine nationale » appartiennent au même champ sémantique pour la majorité des répondants. Le fait d'être interrogé à plusieurs reprises sur ce qu'ils considéraient être un concept identique a incité les répondants à juger bon nombre de questions redondantes. À l'extrême, certains répondants ont cru qu'on leur posait les mêmes questions différemment pour les amener à dévoiler des attitudes racistes cachées.

Phase deux

Consécutivement à la phase un, on a apporté d'importantes modifications au questionnaire du supplément en vue d'aplanir les difficultés relevées lors de la recherche cognitive. La phase deux du projet devait servir à tester le nouvel instrument et comprenait 54 interviews cognitives. Le nombre plus élevé de répondants de chaque race et groupe ethnique a permis l'inclusion de personnes d'une scolarité inférieure au cours secondaire et de personnes cumulant une année ou plus d'études collégiales. Les interviews de la phase deux se sont déroulées aux endroits suivants et portaient sur les groupes raciaux et ethniques que voici : Albuquerque (Amérindiens); Chicago (noirs); Houston (hispaniques, blancs); Nouvelle-Orléans (créoles); New York (hispaniques, blancs); régions rurales de Californie (hispaniques); régions rurales du

Mississippi (noirs); régions rurales de Virginie de l'Ouest (blancs); San Francisco (asiatiques et natifs des îles du Pacifique, hispaniques, multiraciaux).

Les interviews cognitives de la phase deux ont révélé que bon nombre de problèmes cernés à la phase antérieure avaient été réglés. La version remaniée de la question 8a montre comment la révision a contribué à contourner la difficulté.

Q. 8a : Plus tôt, je vous ai lu une liste comprenant les groupes suivants. L'addition d'une catégorie multiraciale à cette liste vous aurait-elle mieux décrit (ou décrit la personne que vous représentez)?

La nouvelle question demandait au répondant s'il aurait préféré se classer ou classer la personne qu'il représente dans la catégorie multiraciale. On a ainsi mis fin aux réponses affirmatives données pour des raisons de « rectitude politique » ou des motifs non pertinents. D'autres questions, révisées elles aussi, ont toutefois continué à soulever des problèmes.

Q 2 : Les gens associent parfois les coutumes, la langue, l'aspect physique ou le pays de naissance à la race. Quel facteur jugez-vous le plus important en ce qui concerne la race d'une personne?

Q 4 : Enfin, quel élément le plus important vous vient à l'esprit quand vous pensez à l'ascendance d'une personne?

Même si les nouvelles questions conceptuelles lui fournissaient des points de repère, le répondant les a trouvées trop abstraites. Ainsi, la question 2 a donné lieu à des réponses du genre : « les gens doivent apprendre à s'entendre » ou « la race n'a pas d'importance ». À la question 4, on a répondu : « les gens devraient connaître leurs antécédents médicaux » ou « il est important de savoir si ses aïeux ont été sauvés ».

Des interviews de la phase deux, il est principalement ressorti que certaines questions restaient confuses ou redondantes pour les répondants, et que le supplément continuait à engendrer une réaction affective négative. On s'est aussi rendu compte qu'il y avait mélange des questions autodescriptives, p. ex. « Quelle est votre race? », et des questions de préférence, p. ex. « Quel terme préférez-vous utiliser pour le groupe racial? », et que cela engendrait une impression de redondance.

Phase trois

On a retravaillé le questionnaire d'après les constatations de la phase deux. Le spécialiste en anthropologie cognitive de l'équipe a suggéré qu'on rende les questions conceptuelles moins abstraites par le truchement d'exemples. Le questionnaire a donc été restructuré pour regrouper les questions de chaque panel en trois parties distinctes : 1) questions d'auto-identification; 2) questions sur les préférences; 3) questions conceptuelles.

La nouvelle version du questionnaire a été testée sur de petits groupes de blancs dans les régions rurales et de personnes hispaniques dans les régions suburbaines. Les répondants des deux groupes ont éprouvé beaucoup moins de difficultés avec la nouvelle version. Ainsi, aucun blanc n'a émis le soupçon qu'on s'efforçait de le leurrer en posant les mêmes questions différemment, comme c'était arrivé antérieurement. D'autres problèmes persistaient toutefois.

Q 2. Les gens associent parfois les coutumes, la langue, l'aspect physique, le pays de naissance ou autre chose à telle ou telle race. D'après vous, quel est l'aspect le plus important dont il faut tenir compte pour déterminer la race de quelqu'un?

Même si on leur demandait d'appliquer le concept de race à une situation particulière, p. ex. caractéristiques ou traits de caractère servant à ranger une personne dans un groupe ethnique quelconque, les répondants ont continué d'éprouver de grandes difficultés avec les questions conceptuelles, même ceux

détenant un diplôme d'études collégiales. Après les interviews de la phase 3, on a donc décidé de supprimer les questions conceptuelles du supplément.

Problèmes non résolus identifiés lors des interviews cognitives

Les interviews cognitives ont fait ressortir d'autres problèmes difficiles à résoudre, notamment le sens que les répondants donnent au terme « multiracial ». Bien que ces derniers aient tous défini le terme par « plusieurs races » ou quelque chose d'approchant à l'interview, ils appliquent difficilement ce terme à eux-mêmes.

Les répondants multiraciaux retenus pour les interviews cognitives avaient des parents de races différentes. Deux répondants identifiés comme « multiraciaux » ne se percevaient toutefois pas comme tels. Ainsi, un jeune homme né de parents hispanique et noir s'est répertorié parmi les « noirs ». Quand on lui a demandé pourquoi, il a déclaré avoir choisi cette catégorie parce qu'on le considérait comme un « noir » dans sa communauté. Un autre, né de parents amérindien et hispanique, s'est spontanément rangé parmi les Amérindiens. Il a subséquemment expliqué qu'il entretenait de piètres relations avec son père et souhaitait qu'on l'associe au même groupe que sa mère. Les personnes que les observateurs classent dans le groupe « multiracial » ne se voient donc pas nécessairement elles-mêmes dans ce groupe.

Les interviews cognitives ont aussi fait ressortir la situation inverse : les personnes que les observateurs avaient classé dans un groupe ethnique précis soutenaient appartenir au groupe « multiracial ». Une femme blanche d'une banlieue de Washington, qui détenait un diplôme d'études collégiales, a choisi cette catégorie en répondant à la question sur la race. Elle a ensuite déclaré qu'elle l'avait fait parce qu'elle était mi-irlandaise, mi-italienne.

De cette dernière et des explications fournies par les répondants quand on les a priés de définir les expressions « race », « ethnicité/groupe ethnique » et « ascendance ou origine nationale », on a conclu que ces concepts se recoupent pour certains non hispaniques et hispaniques. La similitude relative des concepts apparaît également dans les réponses aux questions moins abstraites où le répondant devait indiquer son ascendance ou origine ethnique. Dans quelques cas, on a obtenu la même réponse, c'est-à-dire « blanc », aux questions sur la race et sur l'ascendance ou l'origine ethnique. Certains groupes, implantés aux États-Unis depuis plusieurs générations, dont la population blanche des régions rurales de Virginie de l'Ouest, ne sont pas parvenus à indiquer une ascendance ou une origine nationale, même si quelques personnes se sont souvenues avoir un ancêtre amérindien. (Tous les répondants de cette communauté qui ont passé l'interview avaient des prénoms britanniques.)

2.3. Déroulement de l'enquête

Une fois l'élaboration du questionnaire terminée, on a entrepris d'autres recherches cognitives en vue de faciliter l'interprétation des résultats. Lors de la semaine de collecte des données de la CPS (du 14 mai au 20 mai), des spécialistes en recherche cognitive ont suivi les interviews téléphoniques assistées par ordinateur (ITAO) aux installations de Hagerstown et de Tucson pendant un jour, puis organisé des groupes de réflexion aux deux endroits le lendemain afin d'en apprendre davantage sur l'expérience des intervieweurs. La même semaine, les chercheurs ont accompagné les intervieweurs de la CPS à Tucson et à Miami pour voir comment l'interview se déroulait sur le terrain. On a enregistré 400 ITAO, soit 200 à chaque installation, en prévision d'un codage subséquent du comportement. Les constatations issues de ces travaux ont servi à interpréter les résultats de l'analyse statistique des données recueillies lors du supplément de la CPS.

Contrôle des ITAO

Le contrôle a révélé que l'identification au groupe multiracial peut varier dans l'une ou l'autre direction avec les générations. Une jeune femme du panel 2, par exemple, a rangé sa mère dans le groupe multiracial (noir, amérindien, hispanique), son père parmi les « noirs » et se disait elle-même « noire ». Une mère qui avait indiqué être de race noire, par contre, et d'ascendance amérindienne, a classé sa fille

dans le groupe « multiracial » (noire et amérindienne). Les intervieweurs ont parfois dû expliquer la question sur l'ascendance ou l'origine nationale.

Observation des interviews sur place assistées par ordinateur (IPAO)

Les spécialistes en recherche cognitive ont observé les interviews de la CPS qui se sont déroulées à Miami et à Tucson, en mai, afin d'en apprendre davantage sur la déclaration concernant l'appartenance à la race « hispanique » et au groupe « multiracial ». Ces observations ont fait ressortir des problèmes au sujet de l'interprétation des concepts et de la terminologie appliquée à ces catégories. Voici un résumé de quelques-unes de ces observations.

À Miami, un Mexicain-Américain a choisi « blanc » dans la liste des groupes ethniques du panel 1 et répondu par l'affirmative à la question sur l'utilité d'un groupe multiracial en déclarant « Mexicain-Américain » quand on lui a demandé quels autres groupes devraient s'ajouter à « blanc ». Dans un autre ménage du panel 1, une femme blanche non hispanique mariée à un Mexicain-Américain a déclaré que sa fille était d'origine hispanique, appartenait au groupe multiracial (question sur la race) et aux groupes blanc et hispanique. Dans un troisième cas (un ménage du panel 3), un homme âgé qui avait affirmé être de race blanche et qui avait répondu par la négative à l'option multiraciale, a ajouté « amérindien » quand on l'a interrogé sur son ascendance ou origine ethnique.

Groupes de réflexion avec les intervieweurs de l'ITAO

Les intervieweurs n'ont rapporté aucune difficulté quant à la compréhension de la question sur l'origine hispanique chez les répondants des panels 1 et 2. Plusieurs répondants non hispaniques ont d'abord paru surpris qu'on leur pose cette question. Bref, ils ne s'y attendaient pas. Quelques répondants non hispaniques ont aussi semblé confus ou offusqués qu'on la leur pose en premier, déclarant « Oh non, pas du tout, absolument pas ». Certains répondants hispaniques ont cru qu'on les mettait à part et insistait trop sur leur ascendance. (Quelques répondants non hispaniques ont aussi voulu savoir pourquoi on posait tant de questions sur les personnes d'origine hispanique.)

Codage du comportement face aux questions sur le statut multiracial dans les ITAO enregistrées

Le comportement face aux questions sur le statut multiracial dans les entrevues ITAO enregistrées a été codé pour l'intervieweur et le répondant. Dans le premier cas, on a codé les comportements suivants : écart important ou mineur dans la formulation des questions, vérification des réponses vagues, sondage en cas de questions incomplètes et codage correct ou incorrect de la réponse. Du côté des répondants, on a codé les comportements suivants : demande d'éclaircissements, interruption de la lecture d'une question pour donner la réponse, réponse du genre « ne sais pas » ou refus de répondre, réponse inadéquate ou adéquate, commentaire sur la difficulté ou la nature délicate de la question.

La façon dont l'intervieweur a lu et codé les questions a soulevé relativement peu difficultés outre, à l'occasion, le fait d'escamoter une partie de la liste de races. L'intervieweur n'a pas toujours prononcé correctement le terme « multiracial », le remplaçant parfois par « multicultural » ou « multinational ». Les données du codage révèlent que le répondant ignorait parfois comment signaler son ascendance.

Un des problèmes que soulève l'interview est que le répondant n'a aucune idée de l'enchaînement des questions sur la race et l'ascendance ou l'origine. Il ne peut donc savoir qu'il existe une question distincte sur l'origine ethnique. Certains pourraient donc avoir été incités à inclure l'ethnicité dans leur réponse à la question sur la race. Quelques personnes ont déclaré que tout le monde aimerait une catégorie « multiraciale » afin de signaler ses origines véritables, c.-à-d. Irlandais, Allemand, etc.

3. ACTIVITÉS SUBSÉQUENTES À L'ENQUÊTE

Une fois les données recueillies, on a analysé les réponses aux questions ouvertes sur le statut multiracial et l'ascendance ou l'origine ethnique. Les résultats de cette analyse et les constatations venant de la recherche cognitive antérieure ont permis une meilleure interprétation des résultats de l'analyse statistique des données du supplément, en particulier les observations ambiguës concernant la déclaration du statut multiracial.

Dans le supplément de mai 1995 de la CPS, le groupe multiracial a été ajouté à la question de la race, pour les panels 2 et 4. Dans le panel 2, la question était précédée d'une autre sur l'origine hispanique. Dans le panel 4, une des réponses possibles à la question combinée sur la race et l'origine hispanique était « origine hispanique, latine ou espagnole ».

Le tableau 1 ventile l'identité raciale des personnes qui ont retenu l'option multiraciale dans les panels 2 et 4 (tiré de Tucker et coll., 1996).

Tableau 1. Ventilation du groupe multiracial

	Panel			
	1	2	3	4
	%	%	%	%
Total	-	1,65	-	1,55
Aucune race / NSP / NPR	-	0,02	-	0,00
« Autre » seulement	-	0,51	-	0,22
Une race seulement	-	0,53	-	0,15
BN / NB	-	0,09	-	0,16
Amérind. + 1 race	-	0,20	-	0,28
A/PI + 1 race	-	0,07	-	0,28
1 race + « autre »	-	0,16	-	0,07
2 autres races	-	-	-	0,20
3 races ou plus	-	0,08	-	0,21

On se rend compte que certains répondants qui avaient choisi « multiracial » au panel 2 et au panel 4 n'ont mentionné qu'une race à la question subséquente où il fallait cocher toutes les races applicables. Une préoccupation initiale au niveau de l'analyse statistique était que les interviews de recherche cognitive aient négligé le fait que certains répondants ignorent le sens du mot « multiracial ». La découverte d'un groupe « multiracial d'une race » important nous a amené à diviser celui-ci en « multiracial » (vrai) et « multiracial indéterminé » (faux) pour analyser les données du supplément de la CPS. À cette fin, on a recouru aux définitions que voici :

Multiracial : personne déclarant appartenir à deux ou plusieurs des groupes ethniques énumérés à la question sur la race du panel;

Multiracial indéterminé : personne ne déclarant pas appartenir à deux ou plusieurs des groupes ethniques énumérés à la question sur la race du panel.

Une très grande proportion de répondants du groupe « multiracial d'une race » du tableau 1 ont choisi « autre » comme race, alors que dans le groupe multiracial mentionnant deux races ou plus, une partie des répondants ont choisi un seul groupe ethnique et la catégorie « autre ». Des interviews de recherche cognitive, il ressort que les personnes du groupe multiracial à qui on avait demandé de répondre aux questions sur la race et qui ne retenaient pas le groupe multiracial choisissaient plus souvent la catégorie « autre » avant d'énumérer différents groupes raciaux dans la question ouverte servant de suivi. L'analyse

des réponses libres à la question de suivi posée à ceux qui s'identifiaient à une « autre » race, dans le groupe multiracial, a révélé une vaste gamme d'origines raciales et ethniques.

« Une race = Multiracial, autre »

Les réponses à développement des personnes qui avaient déclaré « 1 race = Multiracial, autre » regroupaient des entrées fort variées du genre créole, Eurasien, Chinois et blanc, Cap-verdien, Allemand et Irlandais. Quelques mentions (créole et Eurasien par exemple), même s'il s'agit de termes simples, représentent en réalité des groupes multiraciaux. Le fait de signaler deux ethnies, p. ex. Allemand et Irlandais, indique qu'il y a chevauchement sémantique des catégories race et ethnicité, comme on l'avait observé lors des interviews de recherche cognitive.

« Deux races = Race + Multiracial, autre »

On a obtenu beaucoup de réponses ouvertes à la question de suivi sur le groupe « multiracial, autre » quand le répondant s'identifiait à une des races sur la liste (p. ex., noir, blanc) et à la catégorie « autre ». Les réponses à la précision « autre » sont suivies de la première lettre de la race choisie parmi celles proposées : Mexicain, Amérindien et Allemand (B); Portoricain, Allemand et Afro-américain (N); Arménien (B); Italien, Néerlandais et Irlandais (B). Le fait d'associer l'ethnicité et la race était donc aussi courant dans ces réponses.

Après l'analyse des précisions à la réponse « autre », il est devenu manifeste que certains membres du groupe « Une race = Multiracial, autre » (p. ex., créole, Chinois et blanc) *cadraient bien* avec la définition du terme « multiracial » élaborée pour l'analyse statistique. On a également constaté que quelques personnes du groupe « Deux races = Race + Multiracial, autre » *ne respectaient pas* la définition, donc devaient faire partie du groupe « multiracial indéterminé ».

Consécutivement à l'analyse des réponses à développement des répondants de type multiracial d'une race et de deux races qui s'identifiaient à la catégorie « autre », on a calculé la proportion de répondants « multiraciaux » (M) et « multiraciaux indéterminés » (MI). Le tableau 2 donne la proportion de ces deux groupes pour les ethnies hispanique et non hispanique. Les personnes qui avaient signalé deux catégories raciales ou plus en dehors de la catégorie « autre » se retrouvent aussi dans le groupe « multiracial ».

Tableau 2. Pourcentage de « multiraciaux » et de « multiraciaux indéterminés »

	Panel 2		Panel 4	
	M	MI	M	MI
	%	%	%	%
Une race nommée				
<i>Hispanique</i>	2,24	10,74	0,0	0,71
<i>Non hispanique</i>	4,88	45,78	5,15	17,02
Deux races ou + nommées				
<i>Hispanique</i>	3,58	4,61	22,79	0,0
<i>Non hispanique</i>	26,03	2,34	52,46	1,88
Total	36,73	63,47	80,40	19,60

La ventilation des groupes multiraciaux indiqués au tableau 1 selon la race révèle qu'environ la moitié des personnes du groupe multiracial qui n'avaient nommé qu'une race pour les deux panels estiment appartenir à un autre groupe racial que le groupe « autre ». Les observations effectuées lors du contrôle des ITAO sur le supplément de la CPS aux installations de Hagerstown donnent une idée des facteurs qui

incitent certaines personnes prétendant appartenir au groupe multiracial à ne déclarer qu'une race. Ces facteurs méritent d'être examinés. On a constaté qu'un répondant disant appartenir au groupe « multiracial » ne mentionnait parfois qu'une race, p. ex. blanc, mais en déclarait une deuxième à la question subséquente sur l'ascendance et l'origine ethnique. La formulation et l'emplacement de cette deuxième question étaient identiques pour tous les panels. La question se lisait comme suit :

À présent, dites-moi quelle votre ascendance ou origine ethnique (ou quelle est l'ascendance ou l'origine ethnique de (nom)).

Après analyse, on constate que 54 % des 152 personnes du groupe « multiracial indéterminé » d'une race qui avaient mentionné une race au lieu de la catégorie « autre » en avaient mentionné une deuxième à la question sur l'ascendance. Pour déterminer l'ampleur du phénomène, on a analysé les réponses de 2 000⁽¹⁾ répondants des panels 2 et 4 qui n'avaient pas mentionné le groupe « multiracial » et n'avaient pas choisi la catégorie « autre ». Ces répondants ont été sélectionnés au hasard, mais proportionnellement à la répartition des races pour les deux panels. Ce groupe a été baptisé « répondants d'une race ». Les réponses de ce groupe à la question sur l'ascendance ne comprenaient une deuxième race, non mentionnée auparavant, que dans 7 % des cas. Ces résultats apparaissent au tableau 3.

Tableau 3. Races supplémentaires mentionnées à la question sur l'ascendance par les membres du groupe « multiracial indéterminé » d'une race et les répondants d'une race

	Races supplémentaires	Aucune race supplémentaire	Total
(non hispanique) « Multiraciaux indéterminés » d'une race	82 (54 %)	70 (46 %)	152
(non hispanique) Répondants d'une race	132 (07 %)	1 750 (94 %)	1 882
Total	214	1 820	2 034

Bien que la taille des deux groupes soit trop faible pour qu'on estime l'importance de l'écart, les données suggèrent que l'existence d'une seconde race parmi les aîeux incite beaucoup de répondants de type « multiracial indéterminé d'une race » à choisir le groupe « multiracial ». Le fait que quelques personnes de deux origines raciales se rangent personnellement dans le groupe « multiracial » alors que d'autres, dans le même cas, se perçoivent comme les membres d'« une seule race » a donné lieu à des recherches démographiques plus poussées sur les deux groupes. (Lire McKay et de la Puente, 1996.)

4. CONCLUSION

Les résultats de la recherche cognitive ont réduit les erreurs de mesure lors de l'élaboration du questionnaire, de la collecte des données à la tenue de l'enquête et de l'interprétation des résultats de l'analyse statistique des données d'enquête. Les chercheurs ont précisé les paramètres des répondants (compréhension des concepts, souhait d'insertion sociale/souci de ne pas révéler des attitudes racistes) et du questionnaire (vocabulaire, effets d'ordre, enchaînement des questions en anglais et en espagnol), ce qui a permis d'en atténuer les effets dans le questionnaire final et de montrer que l'enquête de la CPS se prête mal aux questions sur la race et l'ethnicité.

1. (Dans 118 cas sur les 2 000 répondants d'une race, la réponse à la question sur l'ascendance n'a pu être codée à l'égard d'une deuxième race. Les entrées comprenaient des réponses du genre « Heinz 57 », « Américain » et « un peu de tout ».)

On a atténué les effets dus à l'intervieweur en donnant à celui-ci une formation qui soulignait la nécessité de lire toutes les réponses possibles à la question sur la race, ainsi qu'en recourant à une bande enregistrée spéciale pour apprendre aux intervieweurs parlant l'espagnol à mieux utiliser le questionnaire dans cette langue. Les problèmes comme la prononciation incorrecte du mot « multiracial » par l'intervieweur nous ont aussi amené à envisager un éventuel sous-dénombrement des membres du groupe multiracial, les répondants n'ayant pas entendu le bon terme. Les caractéristiques des répondants relevées durant l'enquête indiquent également la possibilité d'un sous-dénombrement des groupes « multiracial » et « hispanique ». Ces caractéristiques comprennent la tendance à rapporter une deuxième race à la question sur l'ascendance et l'agacement d'avoir à écouter une longue liste de réponses possibles, qui pourrait avoir incité la personne à choisir la nouvelle option « multiracial », venant après l'énumération des groupes habituels. Enfin, les chercheurs ont identifié les effets de mode qui pourraient se présenter lors d'une enquête par interview quand le répondant ne peut deviner qu'on lui posera subséquemment une question sur l'ascendance ou l'origine nationale et précise donc l'ethnie à la question sur la race.

Les résultats de la recherche cognitive ont joué un rôle capital dans l'interprétation des résultats ambigus ou contre-intuitifs issus de l'analyse statistique des données relatives à la déclaration du groupe multiracial. Le fait que les répondants aient confondu race, ethnicité et ascendance à l'interview cognitive a exigé l'analyse des réponses ouvertes de ceux qui avaient déclaré « autre » race pour le groupe multiracial. On a ainsi découvert un groupe « multiracial vrai d'une race » (p. ex., créole, eurasiens) et un groupe « multiracial indéterminé de deux races » (p. ex. blanc et Allemand, noir et afro-américain).

Les résultats des travaux de recherche cognitive entrepris parallèlement à l'élaboration et à l'application du supplément de la CPS sur la race et l'ethnicité ont aussi contribué à réduire l'erreur de mesure associée aux statistiques issues du supplément, y compris le fait que plus de la moitié des répondants de statut multiracial ne répondaient pas à la définition de multiracial, tandis que quelques personnes soutenant ne pas appartenir au groupe multiracial comptaient des aïeux de diverses races.

REMERCIEMENTS

L'auteur tient à remercier les autres membres de l'équipe de recherche cognitive pour leur importante contribution au projet : Adalberto Aguirre (Université de Californie, Riverside), Patricia Bell (Université de l'Oklahoma), Ada Costa-Cash (*Bureau of the Census*), Manuel de la Puente et Eleanor Gerber (*Bureau of the Census*), LuAnn Moy (GAO), Jorge Nakamoto (Aguirre International), et Jaki Stanley (NASS/DOA) ainsi que Clyde Tucker, Brian Kojetin, Shail Butani et Linda Stinson (*Bureau of Labor Statistics*) et Betsy Martin (*Bureau of the Census*).

REMARQUES

Les opinions exprimées dans le présent document n'engagent que l'auteur et ne reflètent pas nécessairement le point de vue du *Bureau of Labor Statistics*.

5. BIBLIOGRAPHIE

BATES, Nancy A, M. DE LA PUENTE, T. J. DE MAIO, et E.A. MARTIN, "Research On Race And Ethnicity: Results From Questionnaire Design Tests," Proceedings of the Census Bureau's 1994 Annual Research Conference, Rosslyn, 107-136.

BENNETT, Claudette, N. MCKENNEY et R. HARRISON, "Racial Classification Issues Concerning Children in Mixed Race Households," Article présenté à la réunion annuelle de la Population Association of America, San Francisco, California, 1995.

BUREAU OF THE CENSUS, Statistical Abstract of the United States: 1993 (113e édition.), Washington, DC, 1993.

DE LA PUENTE, MANUEL et RUTH B. McKAY , "Developing and Testing Race and Ethnic Origin Questions for the Current Population Survey Supplement on Race and Ethnic Origin," Proceedings of the 1995 Annual Meeting of the American Statistical Association, Section on Government Statistics, 19-28.

ELIAS-OLIVARES, L. et M. FARR, Sociolinguistic Analysis of Mexican-American Patterns of Non-Response to Census Questionnaires. Rapport soumis au Census Bureau under Joint Statistical Agreement 88-25, Ethnographic Exploratory Research Report #16, 1991.

KISSAM, Edward, E. HERRERA, et J.M. NAKAMOTO, Hispanic Response to Census Enumeration: Forms and Procedures, Rapport soumis au Census Bureau under Contract No. 50-YABC-2-66027, Task Order No. 46-YABC-2-0001, mars, 1993.

LESSLER, J. T. et M.G. SIRKEN, "Laboratory-Based Research on the Cognitive Aspects of Survey Methodology," Milbank Memorial Fund Quarterly/Health and Society, Vol. 63, 1985, 565-581.

McKAY, RUTH B. et M. DE LA PUENTE, "Cognitive Testing of racial and ethnic origin questions for the CPS Supplement," Monthly Labor Review, Vol. 119, No. 9, 8-12., Sept., 1996.

SNIPP, MATTHEW C., "Who Are American Indians? Some Observations About the Perils and Pitfalls of Data for Race and Ethnicity." Population Research and Policy Review 5: 237-252, 1986.

TUCKER, N. CLYDE, R.B. McKAY, B.A. KOJETIN, R. HARRISON, M. DE LA PUENTE, L. STINSON et E. ROBISON, "Testing Methods Of Collecting Racial And Ethnic Information: Results Of The Current Population Survey Supplement On Race And Ethnicity," Bureau of Labor Statistics Statistical Note Series, No. 40, juin, 1996.

MESURE DE LA QUALITÉ DANS LE TRAITEMENT DES DONNÉES D'ENQUÊTE

Kathryn Williams, Connie Denyes, Mary March, Walter Mudryk¹

RÉSUMÉ

Bien qu'on s'efforce d'éviter les erreurs non dues à l'échantillonnage lors de la planification d'une enquête, des techniques de contrôle de la qualité sont nécessaires pour veiller à ce que les normes qualitatives soient respectées durant l'enquête. Le traitement des données d'enquête intègre bon nombre de procédures complexes où des erreurs sont toujours possibles : les données peuvent être mal saisies ou codées, ou on peut appliquer les méthodes de correction manuelles de la façon inappropriée. L'échantillonnage pour acceptation et l'inspection d'acceptation, le contrôle du processus statistique et l'analyse de Pareto sont des techniques statistiques qui permettent de contrôler ces erreurs en cours de production. Nous décrivons ici comment ces techniques ont été intégrées à un programme de gestion de la qualité du traitement des données d'enquête à la Division des opérations et de l'intégration de Statistique Canada. Des exemples réels d'application du contrôle de la qualité illustreront ce programme structuré de mesure de la qualité et d'améliorations soutenues.

MOTS CLÉS: échantillonnage pour acceptation, inspection d'acceptation, contrôle du processus statistique, normes de qualité, programme de gestion de la qualité, amélioration continue.

1. INTRODUCTION

Statistique Canada attache une grande importance à la qualité, pour de bonnes raisons. Sa réputation en dépend. En guise d'introduction au symposium de 1990, M. Fellegi rappelait que la réputation de Statistique Canada quant à la diffusion de statistiques fiables reposait directement sur l'utilité de ses produits, peu d'utilisateurs ayant la possibilité de répéter les tests ou d'évaluer directement d'une manière quelconque la qualité des produits. C'est pourquoi ils se fient sur la réputation de l'organisme qui les leur propose. Statistique Canada s'est mérité sa réputation en surveillant de près la qualité à toutes les étapes de l'enquête : au moment de décider la teneur de l'enquête ou de réduire les erreurs dues à l'échantillonnage par un meilleur plan, d'augmenter le taux de réponse en recourant à l'interview assistée par ordinateur, d'atténuer l'erreur due au répondant par des corrections ou de diminuer les erreurs introduites lors de la saisie, du codage ou de la correction manuelle des données.

Nous mettrons ici en relief les initiatives prises par la Division des opérations et de l'intégration (DOI) pour garantir la qualité des services de traitement qu'elle dispense dans le cadre des nombreux programmes du Bureau. Ces initiatives sont cohérentes avec les pratiques courantes du secteur privé qui consistent à établir des normes de qualité et à recourir à des méthodes statistiques pour contrôler la qualité et s'assurer que le travail effectué respecte ou surpasse les attentes du client. Ces pratiques sont solidement ancrées dans les principes énoncés par les pionniers des méthodes statistiques de gestion de la qualité, c'est-à-dire Juran (1988), Deming (1986) et Shewhart (1931). Dans les pages qui suivent, nous décrivons les méthodes statistiques de contrôle de la qualité en usage à Statistique Canada et la façon dont on les a intégrées à un programme de gestion de la qualité du traitement. Des exemples illustreront comment ces méthodes et ce programme préviennent ou réduisent le nombre d'erreurs non dues à l'échantillonnage lors du traitement des données d'enquête.

¹Kathryn Williams, Mary March, Walter Mudryk, Division des méthodes d'enquêtes-entreprises, Connie Denyes, Division de la recherche et du développement des opérations, Statistique Canada, Ottawa, Ontario, K1A 0T6.

2. QUESTIONS RELATIVES AUX ERREURS NON DUES À L'ÉCHANTILLONNAGE

Les erreurs non dues à l'échantillonnage qui surviennent lors du traitement résultent habituellement de la collecte, de la saisie, du codage et de la correction des données. On se pose couramment les questions que voici quand vient le moment de s'attaquer à ces erreurs :

- Quelles normes de qualité doit-on respecter?
- De quelle façon mesure-t-on l'erreur non due à l'échantillonnage lors de la saisie des données, de leur codage ou d'autres processus? La mesure a préséance car, comme le disait Deming, on ne peut améliorer que ce qu'on peut mesurer.
- Comment peut-on atténuer ce genre d'erreur?
- Comment peut-on parvenir à des améliorations soutenues? Il est essentiel de surveiller la qualité en permanence et d'assurer la rétroaction.
- Comment cerne-t-on les causes initiales des erreurs et fait-on en sorte qu'elles ne se reproduisent pas?

On peut illustrer ces cinq aspects avec l'exemple véridique que voici. Nous avons relevé un taux d'erreurs plus élevé que la norme de qualité établie pour la saisie des données à l'égard d'un code, dans un questionnaire. Les erreurs ont pu être réduites par inspection et correction. Nous avons ensuite analysé les résultats du contrôle de la qualité et découvert l'origine des erreurs. On s'est ainsi rendu compte que le code à saisir était encerclé en vert, ce que le préposé ne pouvait voir car le questionnaire était de la même couleur. Lors des enquêtes subséquentes, on s'est servi d'un feutre rouge. Une mesure préventive simple, mais efficace.

Le dernier point se rapporte à la hantise du contrôle de la qualité et à la façon d'y mettre fin. Il faut travailler activement pour que chacun comprenne bien que nous ne sommes pas la «police de la qualité». Nous mesurons les erreurs en vue de les repérer et signalons simplement ce qu'il y a lieu d'améliorer.

3. MÉTHODES STATISTIQUES DE CONTRÔLE DE LA QUALITÉ

Les méthodes statistiques de contrôle de la qualité supposent qu'on doit déterminer ce qu'il faut contrôler, établir des normes de qualité, choisir les techniques statistiques appropriées, jauger le rendement réel, analyser l'écart (entre le rendement réel et la norme), puis découvrir et corriger la cause de cet écart.

Les techniques statistiques les plus couramment utilisées à Statistique Canada sont l'échantillonnage pour acceptation, le contrôle du processus statistique et l'analyse de Pareto. On y recourt habituellement dans le cadre de l'inspection d'acceptation (Schilling, 1982). Mains rapports et documents internes parmi lesquels ceux de Mudryk (1988), Mudryk, Croal et Bougie (1994) et de Mudryk, Burgess et Xiao (1996) décrivent comment on a perfectionné et appliqué les instruments de traitement des données d'enquête et comment les systèmes de soutien connexes ont vu le jour.

3.1 Échantillonnage pour acceptation

La théorie sous-jacente à l'échantillonnage pour acceptation revêt une grande importance pour l'inspection d'acceptation. Dans le cadre d'une enquête, l'échantillonnage pour acceptation suppose la division du travail confié à un employé en lots homogènes et la sélection aléatoire d'un échantillon des résultats de ce travail, pour chaque lot, afin de vérifier s'il y a eu des erreurs (et si elles ont été corrigées). Si le nombre total d'erreurs pour un échantillon est inférieur au seuil établi, le lot est accepté; dans le cas contraire, il est rejeté et on entreprend une inspection complète afin de rectifier les erreurs qui subsistent.

La taille de l'échantillon et les critères d'acceptation (n, c) reposent sur une norme de qualité précise. Cette norme, la qualité rectifiée moyenne limite (AOQL), permet de réduire le nombre d'inspections à un taux d'erreurs prévu au départ et suppose un certain degré de risque pour le producteur et le client (Dodge et Romig, 1959). La norme sur laquelle repose le plan d'échantillonnage pour acceptation ne peut être sélectionnée arbitrairement, car plus elle est sévère, plus la probabilité que les lots soient rejetés est grande et plus le taux d'inspection, donc le coût, augmente.

Les figures 1 et 2 illustrent le lien entre le taux d'erreurs d'entrée, le taux d'erreurs de sortie, le nombre d'inspections nécessaires à l'égard d'un plan d'échantillonnage pour acceptation simple (n, c) = (4, 5), pour un lot $N=30$ et une qualité rectifiée moyenne limite (AOQL) de 70 erreurs par centaine d'unités, en supposant une inspection de rectification. Nous choisissons comme exemple la saisie de données d'un questionnaire comportant 200 champs où le taux d'erreurs maximum a été établi à 70 champs pour 100 questionnaires, soit moins de 0,4 % de champs erronés par questionnaire (Duddek, 1996). Si le nombre total de champs erronés de l'échantillon (quatre questionnaires sur 30) est inférieur ou égal au critère d'acceptation ($c=5$), on estime que le lot est acceptable. Si le total dépasse le critère d'acceptation, le lot est rejeté et on devra inspecter tout le lot.

Figure 1

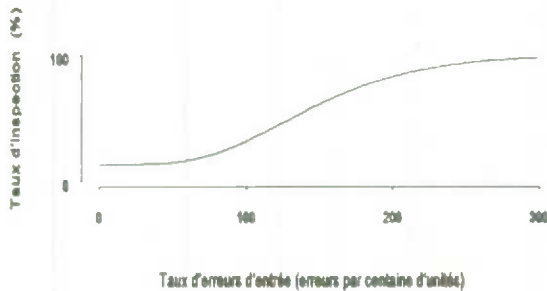
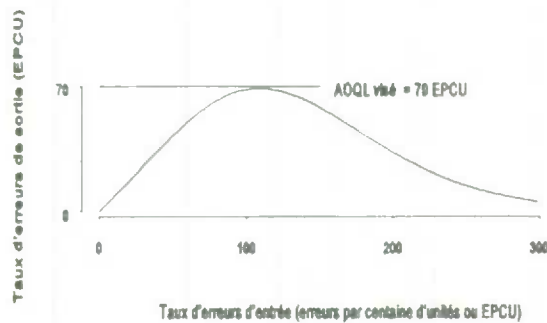


Figure 2



La figure 1 révèle qu'avec la hausse du taux d'erreurs d'entrée, on relève une augmentation du taux d'inspection. On le doit au fait que plus le lot compte d'erreurs, plus le risque qu'il soit rejeté est élevé. La figure 2 montre la courbe suivie par la qualité rectifiée moyenne (AOQ) des lots avec l'inspection. L'inspection de rectification suppose l'élimination des erreurs identifiées par un processus de correction quelconque. Lorsque les erreurs d'entrée augmentent, le nombre d'erreurs de sortie se relève également jusqu'à atteindre un maximum appelé qualité rectifiée moyenne limite (AOQL). Au-delà de ce seuil, le taux moyen d'erreurs rectifiées diminue, car le nombre de lots rejetés et d'erreurs corrigées ne cesse d'augmenter. L'AOQ s'exprime de la façon suivante :

$$AOQ = \frac{P_a p (N-n)}{N}$$

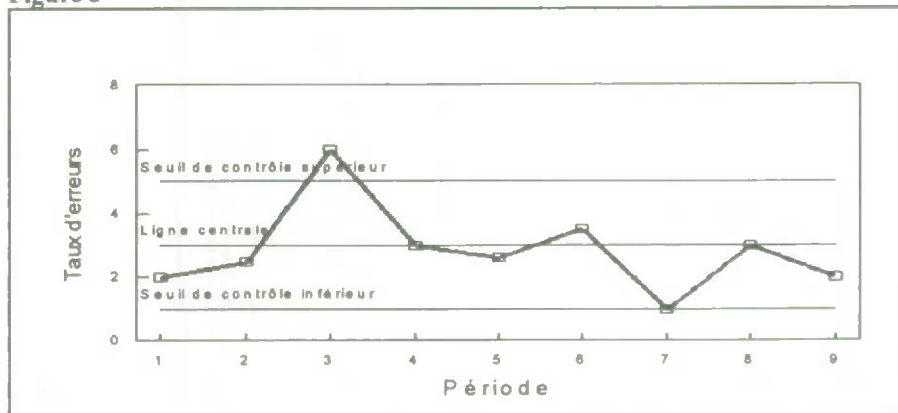
où p représente la fraction d'erreurs d'entrée pour le lot et P_K la probabilité que le nombre d'erreurs d soit inférieur ou égal au critère d'acceptation c . P_K s'exprime comme suit :

$$P_a = P\{d \leq c\} = \sum_{d=0}^c \frac{n!}{d!(n-d)!} p^d (1-p)^{n-d}$$

3.2 Contrôle du processus statistique

Une autre technique statistique qu'on utilise de plus en plus à Statistique Canada est le contrôle du processus statistique (CPS). Cette technique est facile à décrire au moyen d'un graphique reproduisant un cas typique. Ainsi qu'on peut le voir à la figure 3, le graphique de contrôle consiste en une ligne horizontale centrale correspondant à la valeur prévue d'un paramètre, le taux d'erreurs moyen par exemple, et en deux lignes parallèles représentant les seuils de tolérance supérieur et inférieur. Ces deux lignes sont habituellement fixées à trois écarts-types de la ligne centrale. Le graphique peut aussi indiquer les seuils d'avertissement inférieur et supérieur, établis à deux écarts-types de la ligne centrale.

Figure 3



En général, le graphique de contrôle fonctionne comme suit. On prélève des échantillons au hasard à intervalles réguliers pour estimer le taux d'erreurs. On reporte ensuite l'estimation obtenue sur le graphique afin de la situer par rapport à la valeur prévue et aux deux seuils. Si un point dépasse un des seuils de tolérance ou si les données montrent la naissance d'une tendance, on interrompt le processus et on prend des mesures pour découvrir l'origine de la situation «aberrante» et y mettre un terme. Le contrôle du processus statistique permet de gérer objectivement la stabilité du processus d'après les limites statistiques qui apparaissent sur le graphique de contrôle (Shewhart, 1931).

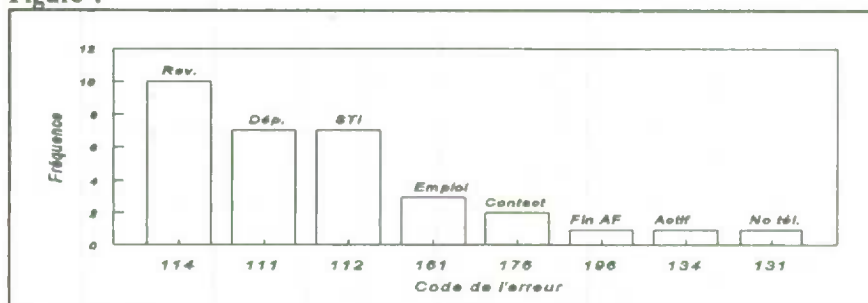
Le contrôle du processus statistique donne les meilleurs résultats quand un processus fonctionne depuis un certain temps et est jugé relativement stable. Cette méthode entraîne des économies car les taux d'échantillonnage sont généralement inférieurs à ceux de l'échantillonnage pour acceptation. Le contrôle du processus statistique n'est pas qu'une technique statistique. En s'engageant à découvrir la cause des erreurs qui surviennent lors de situations «aberrantes», on assure une amélioration continue.

3.3 Analyse de Pareto

Enfin, le graphique de Pareto est un instrument important d'analyse et de rétroaction. Il s'agit d'un cas spécial de distribution des fréquences où les catégories d'erreur se présentent dans l'ordre, de la plus fréquente à la moins fréquente. Cette méthode repose sur le principe de Pareto voulant que la plupart des erreurs résultent habituellement d'un nombre restreint de causes envisageables. Les catégories du graphique correspondent normalement aux différents types d'erreur. Dans la pratique, les premières catégories regroupent souvent une part très sensible des erreurs qui surviennent au début d'une opération. La distribution de Pareto attire l'attention sur ces erreurs, ce qui permet d'analyser rapidement les causes et les effets et de prendre des mesures correctrices sans tarder.

La figure 4 illustre le graphique de Pareto. Cet exemple donne la distribution de Pareto des erreurs de saisie des données pour les champs d'un questionnaire. Le champ «revenu» est celui où les erreurs reviennent le plus fréquemment suivi du champ «dépenses» et du code de la Classification type des industries (CTI). En étudiant l'origine des erreurs relatives à ces trois champs, on réduira sensiblement le nombre global d'erreurs qui résulte de la saisie des données.

Figure 4



3.4 Inspection d'acceptation

Beaucoup d'activités de traitement poursuivies à Statistique Canada sont trop brèves pour engendrer le processus stable nécessaire à un contrôle du processus statistique. C'est pourquoi on recourt à des plans d'échantillonnage pour acceptation, mais dans le cadre d'une inspection d'acceptation. L'inspection d'acceptation suppose une stratégie permanente où les méthodes d'échantillonnage pour acceptation sont sélectionnées et adaptées en fonction des conditions d'inspection (Schilling, 1982). Quand le cycle de vie des activités est plus long, l'instabilité initiale du processus exigera un échantillonnage pour inspection d'acceptation. Avec la stabilisation du processus, on fera appel à des méthodes d'échantillonnage pour acceptation temporaires comme l'inspection d'un échantillon réduit et des plans d'échantillonnage partiels. L'objectif ultime est de mettre fin à l'inspection par échantillonnage pour inspection d'acceptation et de la remplacer par le contrôle du processus statistique au terme d'efforts constants d'amélioration. L'inspection d'acceptation permet de prévenir ou de réduire bon nombre d'erreurs non dues à l'échantillonnage.

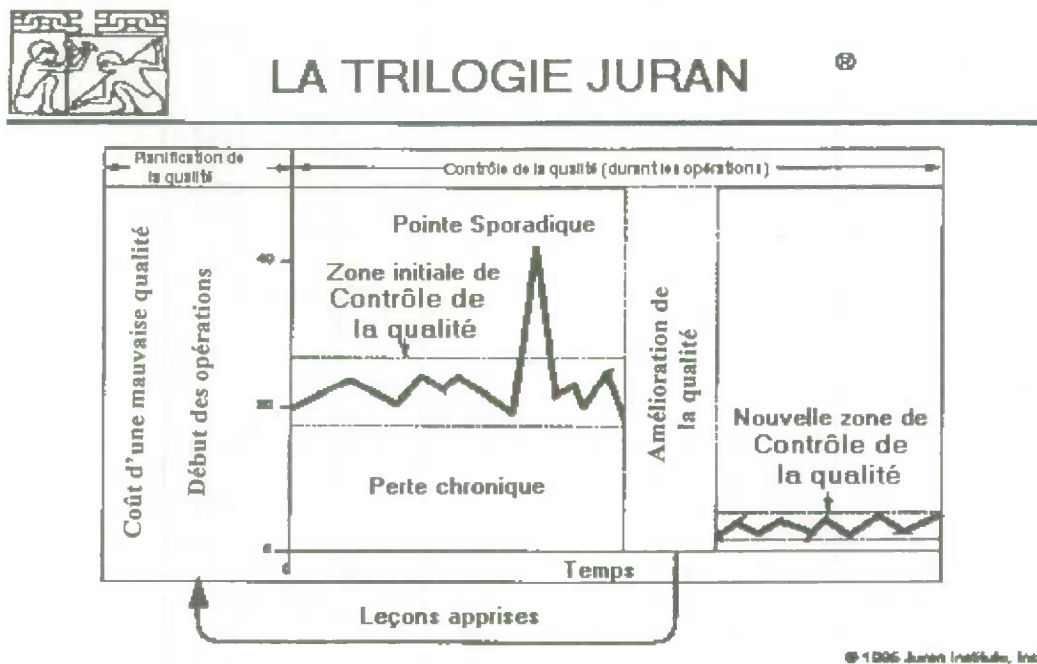
4. PROGRAMME DE GESTION DE LA QUALITÉ

Les méthodes statistiques que nous venons de décrire s'intègrent dans un programme structuré ayant pour but de gérer la qualité au service central de traitement des données d'enquête de Statistique Canada. La Division des opérations et de l'intégration dispense divers services de traitement comme la collecte, la saisie, le codage et la correction des données dans le cadre de nombreuses enquêtes du Bureau. Le programme de gestion de la qualité poursuit plusieurs objectifs :

- faire en sorte que les produits remis à la clientèle respectent ou surpassent les attentes de cette dernière sur le plan de la qualité;
- donner aux cadres de la Division l'assurance que les activités se déroulent comme prévu et qu'on respecte les normes établies;
- améliorer continuellement les activités en relevant la qualité et réduisant les coûts.

Pareil système favorise une amélioration continue des activités d'enquête car il touche la planification, le contrôle et l'amélioration de la qualité. Le programme s'inspire du modèle de gestion de la qualité en usage dans le secteur privé, décrit par Juran (1988) dans son graphique de la «trilogie» en instaurant un cycle permanent de mesure, de rétroaction et d'amélioration. Mudryk (1991) a adapté ce modèle aux activités d'enquêtes et rédigé un document sur le sujet.

Figure 5



La gestion de la qualité suppose trois processus administratifs normalisés : planification, contrôle, amélioration. Comme on peut le constater à la figure 5, la **planification de la qualité** prévoit une conception stratégique des produits et des procédés selon les besoins pressentis et l'établissement d'objectifs en matière de qualité. Dans le graphique servant d'exemple, le coût initial d'une piètre qualité se traduit par la proportion d'unités défectueuses (20 %).

Lorsque la production est lancée, le **contrôle de la qualité** fait en sorte que les objectifs établis à l'étape de planification soient respectés. Les méthodes statistiques comme celles décrites précédemment permettent d'identifier les erreurs ou les carences au niveau de la qualité. Le contrôle de la qualité fait en sorte qu'on règle rapidement toute baisse sporadique de la qualité - comme l'indique le pic de 40 % sur le graphique. Les erreurs restent néanmoins à un niveau stable, ce qui signifie qu'il existe un gaspillage chronique. Cet aspect est associé au coût du reconditionnement des lots rejetés, ou aux erreurs et aux corrections qui surviennent plus tard dans le processus. Le niveau d'erreurs chroniques a souvent son origine dans la conception effectuée durant la période de planification.

L'**amélioration de la qualité** suppose l'examen du taux d'erreurs permanent (gaspillage chronique) en vue d'en déterminer les causes et d'identifier les mesures correctrices durables qui amélioreront le processus à long terme. Au terme des efforts déployés pour améliorer la qualité, le coût de la qualité atteint une nouvelle zone, à 3 %, sur le graphique. Les leçons tirées de ce processus et les causes d'une piètre qualité peuvent servir à planifier la qualité dans l'avenir.

La meilleure façon d'illustrer le programme de gestion de la qualité consiste à suivre une application le long du processus. Tout débute lorsque le service des opérations ou le directeur des enquêtes cerne le besoin d'un contrôle de la qualité. Une équipe composée de membres des services des opérations, des enquêtes et de la méthodologie est réunie et passe en revue les activités en s'attardant à divers aspects comme le processus, les procédures, les contraintes, le volume, les ressources et les échéances. Après avoir déterminé où la qualité sera contrôlée, l'équipe fixe les normes de qualité et détermine la technique statistique à utiliser. Pour établir les normes, on doit équilibrer le degré de qualité désiré avec le coût et les risques. Les méthodes statistiques de contrôle de la qualité auxquelles on recourt à Statistique Canada comprennent habituellement le contrôle du processus statistique (CPS) pour les interviews téléphoniques assistées par ordinateur (ITAO) et l'inspection d'acceptation pour la plupart des autres activités de traitement (à savoir saisie des données, codage, etc.).

Pour instaurer un processus efficace et un bon contrôle de la qualité, il convient d'établir clairement les spécifications du traitement. L'équipe s'assure que ces spécifications existent. Des codes d'erreur sont établis à partir des spécifications et on pondère les différentes sortes d'erreurs advenant le cas où certaines s'avèrent plus graves que d'autres. Après avoir déterminé tous les paramètres et les avoir intégrés aux systèmes de soutien appropriés, on procède aux essais. Cela fait, on rédige les procédures à l'intention des contrôleurs de la qualité, à qui il faut aussi apprendre à repérer les erreurs, à les prendre en note en vue d'une rétroaction et à adopter des mesures correctrices.

Une fois que la production commence, les contrôleurs de la qualité refont une partie du travail ou, dans le cas de l'ITAO, suivent en partie quelques interviews téléphoniques, selon les plans d'échantillonnage préétablis. Les erreurs sont identifiées et corrigées, et s'il s'agit d'erreurs graves, la rétroaction est immédiate ce qui permet de garder la maîtrise du processus. Toutes les erreurs sont consignées et on rédige des rapports à l'intention de l'équipe pour qu'elle puisse les analyser, après avoir rassemblé et résumé les constatations. Les membres de l'équipe cherchent les causes des erreurs et trouvent des mesures correctrices qui auront des effets durables, par exemple des modifications aux processus, aux méthodes, aux procédures, aux manuels ou à la formation. On signale aussi la situation aux responsables du traitement et aux intervieweurs pour qu'ils corrigent certaines erreurs et cherchent à s'améliorer. Enfin, les leçons tirées de l'analyse sont examinées durant la planification, de manière à poursuivre l'amélioration du processus.

Un programme de gestion de la qualité de ce genre présente maints avantages :

- il permet de s'assurer qu'on garde la maîtrise du processus;
- il garantit un certain degré de qualité à l'égard des produits remis aux clients;
- il permet de repérer les erreurs les plus courantes afin que des mesures puissent être prises là où elles seront les plus utiles;
- les résultats et la rétroaction constituent un excellent exercice de formation et d'apprentissage, ce qui débouche sur des employés plus compétents et plus expérimentés dans le traitement des données d'enquête;
- on dispose d'une structure propice à l'amélioration continue du processus, donc à la réduction des coûts.

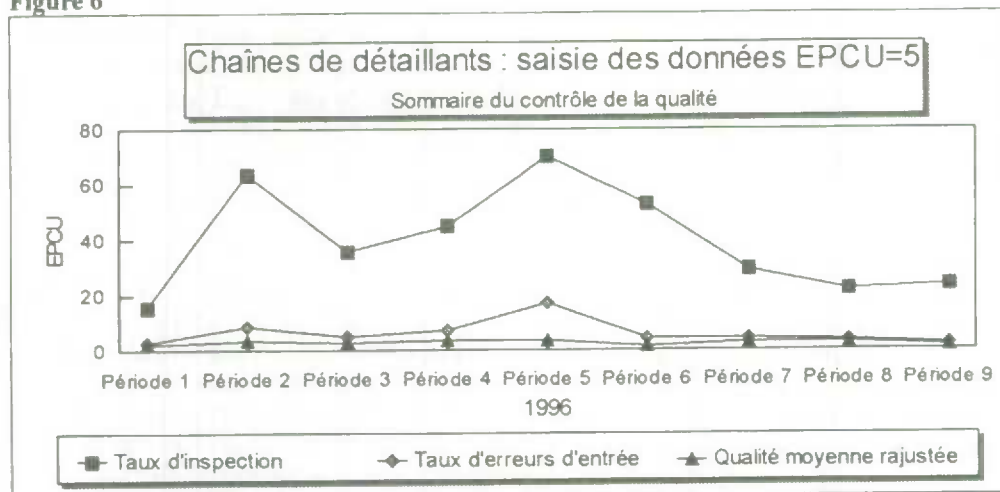
5. APPLICATIONS DU CONTRÔLE DE LA QUALITÉ

La meilleure façon d'illustrer l'application des méthodes statistiques et du programme de gestion de la qualité consiste à prendre des exemples réels de la Division des opérations de l'intégration.

Notre premier exemple concerne l'utilisation du logiciel DC-2 de collecte et de saisie générales des données de Statistique Canada. Ce logiciel, qui comprend un module de contrôle de la qualité, permet de saisir une deuxième fois les données et de les comparer en ligne afin d'en déterminer la qualité. Une enquête qui recourt actuellement à cette technologie est celle tenue chaque année sur les chaînes de magasins de détail. Dans ce cas, on a retenu comme norme de qualité acceptable, cinq erreurs par centaine d'unités (AOQL=5 epcu). En d'autres termes, cinq champs sur 1 000 environ devraient présenter des erreurs pour une centaine de questionnaires. Les plans d'échantillonnage pour acceptation font en sorte que le niveau de qualité rajusté est toujours inférieur à la norme. L'équipe examine les résultats du contrôle de la qualité pour s'assurer que la norme est respectée et pour identifier les améliorations susceptibles d'être apportées au processus de saisie. Résultats et améliorations sont signalés aux personnes qui s'occupent de la saisie des données.

Comme on peut le voir à la figure 6, une hausse du taux d'erreurs d'entrée augmente le taux d'inspection. Dans notre exemple, certaines erreurs sont attribuables au fait que les préposés à la saisie oublient un champ rarement signalé. En examinant les résultats du contrôle de la qualité avec les préposés, on attire leur attention sur les champs qui comportent le plus d'erreurs. La figure montre plusieurs fluctuations du taux d'erreurs d'entrée, signe que le nombre d'erreurs augmente puis baisse lorsqu'on adopte des mesures afin de découvrir l'origine du problème. Dans cet exemple, on a débuté le travail avec un préposé expérimenté. Les deux pics subséquents sont attribuables à l'arrivée de nouveaux employés qui ont été affectés à la saisie des données de l'enquête.

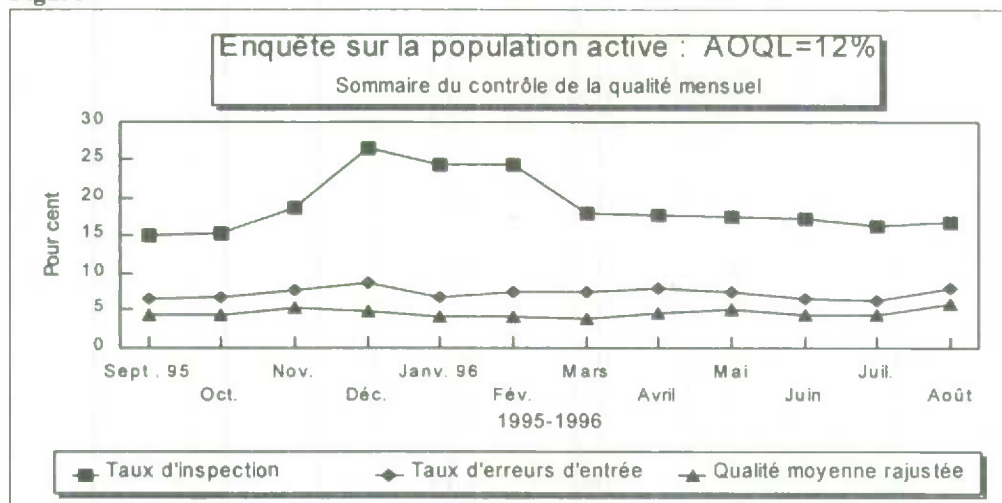
Figure 6



Un deuxième exemple concerne les techniques de contrôle de la qualité instaurés pour l'opération de codage des professions et des branches d'activité de l'Enquête sur la population active. Cette application a été mise au point pour qu'on dispose d'une norme de qualité applicable aux enregistrements qui nécessitent un codage manuel, l'attribution des codes étant une activité aussi subjective que complexe. Dans le cas actuel, la norme de qualité ou AOQL a été fixée à 12 % d'erreurs dans les questionnaires codés à la main. Cette application entièrement automatique du contrôle de la qualité est unique car elle exploite une stratégie articulée sur la règle de la majorité, qui exige une vérification indépendante de deuxième niveau avec troisième codage indépendant lorsque les deux premiers codes diffèrent. Le code retenu correspond au code identique dans deux cas sur trois. Grâce à cette méthode, il est possible de surveiller chaque code de la Classification type des industries (CTI) et(ou) de la Classification type des professions (CTP), donc d'assurer un contrôle de la qualité par code. Les résultats sont communiqués aux préposés afin qu'ils puissent se concentrer sur les codes qui suscitent le plus d'erreurs.

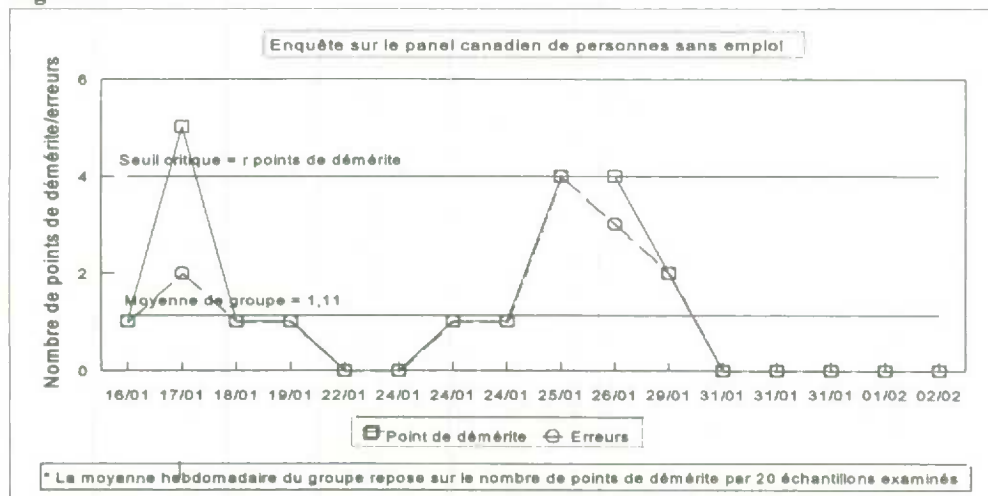
La figure 7 montre que la qualité moyenne rajustée se situe entre 4 % et 6 % pour 12 mois. Ce résultat se trouve 2 à 4 points sous le taux d'erreurs d'entrée et nécessite généralement 15 % à 20 % d'inspection. La réduction du taux d'erreurs de jusqu'à 50 % survenue en mars 1996, par exemple, illustre les avantages du contrôle de la qualité.

Figure 7



Notre troisième exemple se rapporte à l'interview téléphonique assistée par ordinateur (ITAO) pour laquelle on recourt au contrôle du processus statistique (CPS) afin d'évaluer la stabilité du processus en comparant les résultats qualitatifs des entrevues supervisées à un seuil préétabli. La figure 8 donne quelques résultats de l'enquête sur le panel canadien de personnes sans emploi. Le seuil supérieur est établi par l'utilisateur à quatre points de démerite (erreurs pondérées) tandis que le seuil inférieur correspond à zéro. On suit les interviews échantillonnées et les problèmes relevés au niveau de la performance de l'intervieweur, comme les divergences par rapport au scénario, sont notés sous forme de points de démerite. Si le nombre total de points est inférieur au seuil établi, aucune mesure correctrice n'est prise. S'il atteint ou franchit la limite supérieure, en revanche, on signale immédiatement la chose à l'intervieweur afin qu'il puisse corriger ses erreurs. Les commentaires tant positifs que négatifs sont transmis à l'intervieweur, s'il y a lieu. En s'efforçant d'aplanir les difficultés dès le départ, on s'assure que chaque activité comme ITAO connaîtra des améliorations continues.

Figure 8



Les exemples qui précèdent ne donnent qu'un petit aperçu des nombreuses applications du contrôle de la qualité à la Division des opérations et de l'intégration.

Un dernier exemple de contrôle du processus statistique mérite d'être mentionné. On a entrepris l'élaboration d'une application du contrôle de la qualité pour le système automatique de saisie des données (SASD), qui recourt à l'imagerie électronique et à la reconnaissance intelligente des caractères. Une fois mise en oeuvre, cette application permettra la surveillance et la correction des erreurs mécaniques et humaines quand le taux d'erreurs dépasse les seuils établis. Le perfectionnement et la promotion de ce système innovateur et rentable exige qu'on mesure et qu'on contrôle la qualité des extraits.

6. CONCLUSION

Nous avons montré de quelle façon les méthodes statistiques de contrôle de la qualité et le programme de gestion de la qualité réussissent à prévenir ou à réduire les erreurs non dues à l'échantillonnage. Deming soutenait que pour avoir du sens, une norme (qu'elle s'applique au rendement, à la sécurité ou à la capacité) doit être définie en termes statistiques. L'élaboration d'une enquête quelconque devrait inclure l'établissement de telles normes, ainsi que le contrôle et l'amélioration du rendement en fonction de cette norme, par le biais de méthodes statistiques. L'examen de la nécessité d'un contrôle de la qualité devrait devenir partie intégrante du développement d'une enquête au même titre que l'examen de la nécessité de la correction des erreurs. Les méthodes de ce genre peuvent s'appliquer aux procédés automatisés (ITAO, SASD) ou manuels. Le cadre structuré servant à gérer et à améliorer constamment le contrôle de la qualité que nous venons de voir pourrait compléter les exercices de restructuration et de reconception entrepris dans le cadre du traitement des données d'enquête.

Les bonnes pratiques de la Division des opérations et de l'intégration ne font pas seulement en sorte que les normes de qualité acceptées par la clientèle soient respectées, elles concourent à bâtir la réputation de Statistique Canada.

7. BIBLIOGRAPHIE

- Deming, W.E. (1986), *Out of the Crisis*, Cambridge, MA: MIT Press.
- Dodge F., Romig H., (1959) *Sampling Inspection Tables*, 2nd ed., John Wiley & Sons, New York.
- Duddek, C. (1996), *Modelling Quality Control Strategies for the 1996 Census of Agriculture*, 1996 ASA Proceedings.
- Early, J.F. (1990), *La gestion de la qualité dans les programmes statistiques nationaux*, Actes du symposium 1990 de Statistique Canada, Mesure et amélioration de la qualité des données, Ministère de l'Industrie, des Sciences et de la Technologie.
- Fellegi, I. (1990), *Mot d'ouverture*, Actes du symposium 1990 de Statistique Canada, Mesure et amélioration de la qualité des données, Ministère de l'Industrie, des Sciences et de la Technologie.
- Juran, J.M. (1988), *Juran's Quality Control Handbook*, 4th ed., McGraw-Hill, New York.
- Montgomery, D. C., *Introduction to Statistical Quality Control*, 2nd ed., John Wiley & Sons.
- Mudryk, W. (1988), *Système de gestion de la qualité dans les opérations d'enquêtes*, Techniques d'enquête, Vol. 14, No. 2, décembre 1988, pages: 327-334, Statistique Canada.
- Mudryk, W., Burgess, M.J., Xiao, P. (1996), *Quality Control of CATI Operations in Statistique Canada*, 1996 ASA Proceedings, Section on Survey Research Methods, Chicago, Illinois.
- Mudryk, W., Croal, J., Bougie, B. (1994), *Generalized Data Collection and Capture (DC-2) Release 2.5.1, Sample Verification (SV) Quality Control Methodology Manual*, Statistique Canada.

Mudryk, W. (1991), *Quality Management: A Framework*, Rapport interne, Statistique Canada.

Schilling, E. (1982), *Acceptance Sampling in Quality Control*, Chapitre 19, Marcel Dekker Inc., New York.

Shewhart, W.A. (1931), *Economic Control of Quality of Manufacturing Processes*, ASQC, Milwaukee, WI.

CONTRÔLE DES INTERVIEWS TÉLÉPHONIQUES ASSISTÉES PAR ORDINATEUR AU *U.S. BUREAU OF THE CENSUS*

Mary Ellen Beach, Jane Woods et Geraldine Burt¹

RÉSUMÉ

Le U.S. Bureau of the Census a toujours attaché la plus haute importance à la production de données de qualité, tout en cherchant à obtenir des taux de réponse élevés dans ses enquêtes. Le contrôle des intervieweurs est une des principales méthodes d'assurance de la qualité dans les centres téléphoniques du Bureau. Il permet d'observer les comportements des intervieweurs qui contribuent aux erreurs non dues à l'échantillonnage, qui découragent la coopération des répondants et qui empêchent la collecte de données exactes et complètes. Le contrôle des intervieweurs a pour but d'assurer que : 1) les directives prescrites concernant les enquêtes soient respectées; 2) les méthodes d'interviews prescrites soient observées dans le respect des règles de l'étiquette et 3) l'instrument d'enquête fonctionne conformément aux exigences et aux besoins du client. Le présent article décrit le système de contrôle mis en place par les trois centres d'interviews téléphoniques assistés par ordinateur (ITAO) exploités par le U.S. Bureau of the Census. Le système actuel est le fruit d'une évolution d'une dizaine d'années; il fournit à la fois une rétroaction immédiate sur le comportement à l'interview et une note qui devient partie intégrante du système officiel de mesure de la performance.

MOTS CLÉS: Contrôle; ITAO; intervieweurs; assurance de la qualité; rétroaction; évaluation.

1. INTRODUCTION

1.1. Aperçu général

Le U.S. Bureau of the Census gère trois centres d'interviews téléphoniques assistés par ordinateur (ITAO) qui emploient chacun de 80 à 110 intervieweurs et qui recueillent des données pour une vaste gamme d'enquêtes démographiques et économiques. Le Centre téléphonique de Hagerstown (HTC), à Hagerstown (Maryland), à environ 75 milles de Washington, a ouvert ses portes en 1985. Un deuxième Centre ITAO a été inauguré à Tucson (Arizona) en 1992. Le Centre téléphonique de Tucson (TTC) est en mesure de conduire des interviews en langue espagnole en plus d'offrir une meilleure couverture des fuseaux horaires et un service de reprise en cas de catastrophe, de jouer un rôle de suppléance et de permettre une plus grande productivité au réseau. En 1994, le Bureau a mis sur pied un service d'interviews assistés par ordinateur (IAO) « multifonctionnel » à Jeffersonville (Indiana). Le Centre téléphonique de Jeffersonville (JTC) a été conçu pour accroître le rendement du ITAO et tirer parti des technologies de l'interview assistée par ordinateur telles que l'entrée des données assistée par ordinateur (CADE), l'entrée des données à l'aide d'un téléphone à clavier (TDE) et l'entrée des données à l'aide d'une unité de reconnaissance de la voix (VRE). Les ressources ainsi consacrées au nouveau matériel et au nouveau personnel pouvaient être partagées par divers clients utilisant la plus grande gamme disponible de technologies.

Selon l'enquête à réaliser, le ITAO est en mesure 1) de partager la charge de travail avec les équipes d'intervieweurs sur le terrain; 2) d'offrir un service peu coûteux de suivi des non-réponses pour les enquêtes ou les recensements envoi/retour par la poste ou 3) de devenir l'unique outil de collecte des données. Les Centres ITAO comptent actuellement un total de 250 postes de travail.

1.2. Aspects techniques

Les intervieweurs des trois centres réalisent toutes sortes d'enquêtes à l'aide des logiciels MicroCATI ou CASES. Les enquêtes effectuées depuis longtemps par le ITAO comme le *Current Population Survey* (données sur la

¹Mary Ellen Beach, Geraldine Burt, *U.S. Bureau of the Census*, Washington, DC 20223; Jane Woods, *U.S. Bureau of the Census*, Jeffersonville, IN 47132.

population active), le *National Crime Victimization Survey* (données sur la criminalité) et le *Quarterly Apparel Survey* (données sur l'industrie) utilisent le logiciel MicroCATI. Les enquêtes dont l'origine est plus récente telles que le *National Survey of Fishing, Hunting, and Wildlife Associated Recreation* (données sur les activités de loisirs et de sports), le *American Community Survey* (recensement détaillé) ainsi qu'une enquête portant sur les employeurs ont été élaborées avec l'aide du logiciel CASES. Le logiciel MicroCATI a été mis au point et est géré par le *U.S. Census Bureau*. Il est graduellement remplacé par le logiciel CASES (*Computer-Assisted Survey Execution System*), un produit commercial mis au point par la *University of California* à Berkeley.

2. CONTRÔLE DU ITAO AU BUREAU

2.1. Pourquoi contrôler?

L'erreur induite par l'intervieweur est une des principales sources d'erreurs non dues à l'échantillonnage commises lors des interviews personnelles et des interviews téléphoniques. Aux trois centres ITAO du *Census Bureau*, le contrôle des intervieweurs sert à dépister et à prévenir ces erreurs. Il permet : 1) d'évaluer l'aptitude d'un intervieweur à bien saisir les concepts qui sous-tendent l'enquête et à appliquer les procédures prescrites pendant son travail; 2) à évaluer le rendement au travail des nouveaux intervieweurs et 3) à déterminer des moyens de réduire les erreurs non dues à l'échantillonnage en améliorant les méthodes d'enquête. L'intervieweur peut faire l'objet d'un contrôle à tout moment, et chacun des centres s'efforce de contrôler environ 2,5 à 5,0 % du temps d'interviews. Lorsqu'un intervieweur est chargé de plusieurs enquêtes, on cherche à contrôler son travail pour chacune de ses enquêtes.

2.2. Qui sont les responsables des contrôles?

Dans les trois centres téléphoniques du Bureau, les contrôles sont réalisés par des surveillants ou par des « moniteurs » spécialement formés à cette fin. Deux des centres ont recours à des surveillants tandis qu'un troisième recrute et forme des « moniteurs ». Les façons de procéder de chaque centre sont un reflet de la philosophie qui sous-tend les contrôles dans ces centres. Le centre de Jeffersonville (Indiana) cherche à atténuer l'aspect « surveillance » de l'exercice pour promouvoir plutôt l'aspect « encadrement ». On espère que les intervieweurs seront plus réceptifs à la rétroaction constructive si la personne chargée du contrôle n'est pas un supérieur hiérarchique direct. Personne n'a cherché à déterminer laquelle des méthodes utilisées donnait les meilleurs résultats. Toutefois, au moins un des autres centres envisage de mettre en place ce système d'encadrement.

2.3. Quand les contrôles sont-ils réalisés?

Le Bureau a recours à trois types de contrôle : contrôle initial, contrôle systématique et contrôle spécial. Le choix du type de contrôle dépend de l'expérience ou du rendement actuel de l'intervieweur.

On a recours au contrôle INITIAL pendant les trois premières séances de travail des intervieweurs nouvellement recrutés. Ce contrôle sert de transition entre la formation en classe et le travail en conditions réelles pour les enquêtes régulières. L'intervieweur n'est pas autorisé à poursuivre l'enquête lorsqu'il n'obtient pas une note suffisante au contrôle initial. En outre, tous les intervieweurs expérimentés qui ont déjà terminé avec succès leurs séances de contrôle initial doivent en subir une nouvelle pour chaque nouvelle enquête confiée à leur centre.

Le contrôle SYSTÉMATIQUE est celui auquel sont soumis tous les intervieweurs qui ont réussi avec succès leur contrôle initial. L'intervieweur restera sous le régime des contrôles systématiques à moins de recevoir une note moins que satisfaisante lors d'une session particulière ou de fournir un rendement insatisfaisant pour une des catégories prévues du formulaire de contrôle. Le contrôle systématique sert historiquement d'outil d'encadrement et de formation. Plus récemment, on a approuvé son utilisation pour l'évaluation du rendement.

Les contrôles SPÉCIAUX sont normalement utilisés en cas de problème de rendement. Les intervieweurs qui éprouvent des difficultés sont contrôlés plus fréquemment, bénéficient d'une rétroaction exhaustive et reçoivent, le cas échéant, une nouvelle formation. Les intervieweurs en difficulté, comme les nouvelles recrues, doivent

réaliser un minimum de trois sessions satisfaisantes consécutives au cours d'une période d'enquête pour être réintégrés sous le régime des contrôles systématiques. Lorsque l'intervieweur ne réussit pas à s'améliorer après une nouvelle formation et un encadrement serré, il est retiré du personnel d'enquêtes.

2.4. Quels sont les aspects contrôlés?

Le formulaire de contrôle comporte six grandes sections :

1. **VOIX ET FAÇON DE S'EXPRIMER ...** Le moniteur/surveillant écoute l'intervieweur pendant son travail afin de s'assurer qu'il s'exprime d'une façon professionnelle, avec une voix agréable et un débit approprié. L'intervieweur doit également s'adapter au débit du répondant, se montrer amical, articulé, poli et convaincant.
2. **APTITUDES À LA LECTURE ...** Les questions doivent être posées telles qu'elles sont énoncées dans les instruments d'enquêtes. Il s'agit là d'une exigence importante pour toutes les enquêtes, qui vise à éviter le biais dû à l'interprétation donnée par les intervieweurs à des questions particulières. Le moniteur/surveillant veille à ce que toutes les questions soient lues telles qu'elles sont énoncées, sans aucune omission. Il vérifie également si l'intervieweur met l'accent approprié sur certains mots précis et vérifie les réponses données comme l'exige le plan d'enquête.
3. **APTITUDES À APPROFONDIR ...** Les intervieweurs doivent savoir reconnaître les réponses incomplètes et obtenir les éclaircissements appropriés. Ils doivent souvent pour cela interroger le répondant en pesant soigneusement leurs mots pour éviter d'influer sur la réponse dans un sens ou dans l'autre. Le moniteur/surveillant écoute ces demandes de précisions pour s'assurer qu'elles sont appropriées.
4. **ENTRÉE DES RÉPONSES ...** Le moniteur/surveillant examine (à l'écran) la façon dont l'intervieweur classe les réponses à chaque question et entre les données. L'intervieweur devrait avoir une maîtrise suffisante de l'instrument d'enquête pour s'y déplacer facilement et détecter les problèmes. En outre, il doit toujours entrer des notes claires et concises dans la section d'informations de l'instrument afin qu'un autre intervieweur éventuellement chargé de finir le travail ou de fermer le dossier puisse se préparer convenablement.
5. **TRAITEMENT DES CAS PROBLÈMES ...** L'intervieweur doit être prêt à traiter efficacement avec les répondants difficiles. Le moniteur/surveillant note la façon dont il se débrouille en cas de problèmes. L'intervieweur doit être capable de convaincre les répondants réticents, de contourner les barrières linguistiques, de traiter avec les répondants hostiles, de garder son calme et de maintenir un ton professionnel.
6. **MAÎTRISE DES CONCEPTS ...** L'intervieweur devrait être en mesure de se présenter, de présenter l'enquête et de répondre aux questions des répondants concernant cette enquête. Le moniteur/surveillant écoute la conversation et détermine dans quelle mesure l'intervieweur maîtrise les concepts de l'enquête et arrive à répondre correctement et avec assurance aux questions qui lui sont posées.

2.5 Méthodes de contrôle

Jusqu'à récemment, même s'il existait un certain consensus quant aux aptitudes particulières à contrôler chez les intervieweurs (voix et façon de s'exprimer, aptitudes à la lecture, etc.), chaque centre téléphonique prenait ses propres décisions concernant la gestion et l'utilisation des résultats des contrôles. Aujourd'hui, les trois centres utilisent le même système de notation pour les opérations de contrôle (S+, S, S- pour le contrôle initial, et O, C, FS, M et U pour les contrôles systématiques et les contrôles spéciaux). La fréquence des contrôles est aussi essentiellement la même aux trois centres.

On peut compter à tout moment jusqu'à sept moniteurs/surveillants à l'œuvre dans chacun des centres. Ces personnes regardent et écoutent les interviews à l'aide d'écouteurs téléphoniques et d'un ordinateur personnel. À chaque poste de contrôle, un système d'affichage permet au contrôleur d'identifier le poste de travail de

l'intervieweur sélectionné. Il n'a ensuite qu'à composer le numéro du poste voulu pour écouter le déroulement de l'interview. Le contrôleur entend l'intervieweur et le répondant. Il peut également surveiller ce que l'intervieweur voit à son écran. Il peut ainsi évaluer le rendement de l'intervieweur, les réactions du répondant, l'utilisation de l'instrument d'enquête et l'entrée des données dans cet instrument. Le contrôleur peut entrer lui-même à l'écran de contrôle les données sur le déroulement de l'interview qui lui permettront plus tard d'assurer la rétroaction. Chaque intervieweur peut être surveillé simultanément par plusieurs contrôleurs. Le contrôle peut être effectué sur place, à partir de « postes » de contrôle installés au service du contrôle, ou à distance, à partir de l'administration centrale de recensement de Suitland (Maryland).

L'application uniforme des procédures de contrôle et des méthodes d'évaluation subjectives de chacune des six catégories susmentionnées est primordiale. Périodiquement, les moniteurs/surveillants se réunissent pour procéder au contrôle d'un intervieweur à l'aide d'un appareil pour conférences téléphoniques. Ils notent leurs observations et leur évaluation comme ils le feraient normalement (sur papier). Une réunion est ensuite organisée pour examiner l'évaluation de chacun des participants. On relève ainsi rapidement les incohérences, ce qui permet aux moniteurs/surveillants de s'entendre sur la façon de les résoudre. On cherche ainsi à assurer l'uniformité de la rétroaction pour les intervieweurs, peu importe la personne chargée du contrôle.

2.6 Méthodes de rétroaction

Les contrôles réalisés aux fins de l'assurance de la qualité et de l'évaluation du rendement seraient inutiles sans rétroaction. En outre, cette rétroaction doit être assurée en temps utile, en particulier lorsqu'elle porte sur des enquêtes périodiques ou ponctuelles. Les intervieweurs qui connaissent une session supérieure à la moyenne ou moins que satisfaisante obtiennent une évaluation officielle. Dans la mesure du possible, la rétroaction est fournie immédiatement après la séance de contrôle lors d'une entrevue privée. On a recours à une évaluation officieuse pour les séances de contrôle dont les résultats ne s'écartent pas sensiblement du rendement moyen attendu. Dans le cas des contrôles systématiques, la rétroaction est souvent fournie au poste de travail de l'intervieweur puisque la quantité d'informations à communiquer est limitée. Toutefois, les personnes qui font l'objet d'un contrôle initial reçoivent toujours une rétroaction officielle puisque se sont de nouvelles recrues.

Pour la séance de rétroaction, le moniteur/surveillant a recours au principe dit du « sandwich ». Il s'agit de commencer la séance en soulignant les aspects positifs du travail, puis d'aborder les points qui méritent d'être améliorés, pour conclure finalement la séance sur une note positive. Cette méthode permet à l'intervieweur de conserver une certaine mesure d'estime de soi et devrait l'encourager à améliorer son travail. La rétroaction positive au début et à la fin de la séance d'évaluation incite également l'intervieweur à se montrer plus réceptif aux commentaires du moniteur/surveillant concernant les aspects à améliorer.

Pendant la séance de rétroaction, le moniteur/surveillant examine la copie papier de la fiche de contrôle avec l'intervieweur. Il peut relever des points et des aspects précis pour décrire en détail les aspects positifs et les points à améliorer. À la fin de la séance de rétroaction, le moniteur/surveillant et l'intervieweur signent le formulaire qui est ensuite entré dans une base de données. La copie papier est ensuite classée. Les intervieweurs peuvent obtenir sur demande des copies de leurs évaluations.

3. ANALYSE DES RÉSULTATS

Pour les besoins du présent article, nous avons examiné les données de contrôle portant sur le *Current Population Survey* (CPS) provenant de deux des trois centres téléphoniques (Tucson et Jeffersonville) pour la période de novembre 1995 à mai 1996.² Le CPS est une enquête nationale portant sur la population active. Il s'agit d'une

²Entre le 1^{er} novembre 1995 et le 30 janvier 1996, un grand nombre de fonctionnaires du gouvernement américain ont pris congé. Le nombre d'interviews et de contrôles réalisés pendant cette période a donc été réduit.

enquête mensuelle ITAO/IPAO³ qui s'étend généralement du dimanche au mardi de la semaine suivante. Sauf en décembre, le CPS est toujours réalisé pendant la semaine qui contient le 19^e jour du mois. Aux centres ITAO, on met habituellement un terme au CPS le mercredi de la première semaine pour retourner les cas recyclés aux responsables sur le terrain. La plupart des interviews ITAO du CPS sont réalisées au cours des deux premiers jours de la période d'enquête. La durée moyenne des interviews est inférieure à 10 minutes. Les interviews mensuelles supplémentaires varient en longueur et ajoutent à la durée moyenne de l'interview. L'intervieweur moyen consacre 20 heures ou moins par mois au CPS. En règle générale, chacun des intervieweurs participant au CPS doit être soumis à un contrôle de 30 minutes pendant la période d'enquête, ce qui équivaut à un pourcentage oscillant entre 2,5 et 5,0 % des heures d'interviews réalisées.

TABLEAU 3.1. FRÉQUENCE DES SÉANCES DE CONTRÔLE PAR TYPE DE SÉANCE

	TTC			JTC		
	% INIT.	% SYST.	% SPÉC.	% INIT.	% SYST.	% SPÉC.
Novembre 1995	73,1	26,9	0,0	35,7	64,3	0,0
Décembre 1995	28,6	68,3	3,2	40,0	60,0	0,0
Janvier 1996	22,5	74,2	3,4	21,4	78,6	0,0
Février 1996	39,8	59,3	0,9	54,8	45,2	0,0
Mars 1996	17,5	79,0	3,5	51,1	48,9	0,0
Avril 1996	5,3	91,7	3,0	42,1	57,9	0,0
Mai 1996	6,3	91,8	1,9	2,9	97,1	0,0

3.1 Fréquence des contrôles

Nous présentons au tableau 3.1 les données de fréquences relatives des trois types de séances de contrôle, pour chaque mois, dans le cas des enquêtes du CPS réalisées à partir des centres téléphoniques de Tucson et de Jeffersonville (TTC et JTC respectivement). Les tests de chi carré révèlent l'existence d'une corrélation statistiquement significative entre l'établissement et la fréquence d'ensemble du type de séance de contrôle. Pour la période de novembre 1995 à mai 1996, on observe également une corrélation claire entre le mois et la proportion des différents types de contrôles réalisés dans le cadre du CPS, tant à Tucson qu'à Jeffersonville. À Jeffersonville, pendant la période à l'étude, les séances de contrôle initial réalisées dans le cadre du CPS constituaient une portion beaucoup plus grande de l'ensemble des séances de contrôle. Le centre de Jeffersonville procédait en effet à cette époque à la formation de 50 nouveaux intervieweurs sur le CPS de février, en prévision du CPS de mars. Les 50 nouveaux intervieweurs ont été soumis au contrôle initial en février et en mars 1996.

3.2. Résultats des contrôles

Les tableaux 3.2a à 3.2f montrent le classement des intervieweurs qui ont subi les contrôles initiaux dans chacun des centres. S+ signifie excellent, S signifie bon à très bon et S- signifie que des améliorations sont souhaitables. Il n'y a pas de différence statistiquement significative⁴ entre le pourcentage des intervieweurs ayant reçu une note satisfaisante ou insatisfaisante dans aucune des catégories de contrôle (p. ex., voix et façon de s'exprimer, aptitudes à la lecture) au sein de chaque centre ou d'un centre à l'autre. À Tucson, les résultats des contrôles initiaux pour le CPS montrent que les nouveaux intervieweurs sont plus nombreux à devoir améliorer leur lecture des questions,

³IPAO est l'acronyme qui désigne les interviews personnelles assistées par ordinateur. Au Bureau, les IPAO sont réalisées soit au domicile des répondants, soit par téléphone, à partir du domicile de l'intervieweur qui utilise à cette fin un ordinateur portatif.

⁴Les tests de signification statistique ont été effectués au seuil de probabilité de 0,05.

leurs aptitudes à approfondir et leur maîtrise des concepts. À Jeffersonville, les résultats des contrôles initiaux montrent que les nouveaux intervieweurs sont plus nombreux à devoir améliorer leur maîtrise des concepts de l'enquête, leur travail d'entrée des réponses et leurs aptitudes à approfondir. Ces résultats portent à conclure qu'on devrait apporter une attention particulière aux aptitudes à approfondir, à la maîtrise des concepts et aux aptitudes à la lecture pendant l'ensemble des séances de formation initiale, tant dans un centre que dans l'autre.

TABLEAU 3.2a. VOIX/FAÇON DE S'EXPRIMER

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=129	JTC n=80
S+	3,9	1,3
S	93,8	93,8
S-	2,3	5,0

TABLEAU 3.2b. APTITUDES À LA LECTURE

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=127	JTC n=80
S+	6,3	2,5
S	75,6	82,5
S-	18,1	15,0

TABLEAU 3.2c. APTITUDES À APPROFONDIR

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=79	JTC n=74
S+	2,5	0,0
S	82,3	86,5
S-	15,2	13,5

TABLEAU 3.2d. ENTRÉE DES RÉPONSES

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=135	JTC n=80
S+	4,4	0,0
S	85,2	83,8
S-	10,4	16,3

TABLEAU 3.2e. CAS PROBLÈMES

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=12	JTC n=28
S+	33,3	0,0
S	66,7	92,9
S-	0,0	7,1

TABLEAU 3.2f. MAÎTRISE DES CONCEPTS

CONTRÔLE INITIAL	% DES INTERVIEWEURS NOTÉS	
	TTC n=98	JTC n=73
S+	3,1	0,0
S	84,7	79,5
S-	12,2	20,5

Les tableaux 3.3a à 3.3f présentent le classement des intervieweurs aux contrôles systématiques à chacun des centres ITAO. O signifie exceptionnel, C digne de mention, FS entièrement satisfaisant, M faible et U insatisfaisant. Il existe une corrélation statistiquement significative entre les notes données aux aptitudes à approfondir, à l'entrée des réponses et à la maîtrise des concepts et de l'instrument. Les résultats portent à conclure que la lecture des questions telles qu'elles sont énoncées constitue le plus grave problème relevé lors des contrôles systématiques réalisés à Tucson. À Jeffersonville, les problèmes principaux concernent l'entrée des données, les aptitudes à approfondir et la lecture des questions. Tel que mentionné plus tôt, on a procédé à Jeffersonville, pendant cette période, à la formation de plus de 50 nouveaux intervieweurs et l'entrée des questions semble présenter un grave problème (11,7 % des intervieweurs notés faibles ou insatisfaisants) parce que les nouvelles recrues prennent un certain temps pour se familiariser avec leur ordinateur.

TABLEAU 3.3a. VOIX/FAÇON DE S'EXPRIMER

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=560	JTC n=137
O	3,0	3,6
C	13,0	6,6
FS	82,1	88,3
M	1,1	1,5
U	0,7	0,0

TABLEAU 3.3c. APTITUDES À APPROFONDIR

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=408	JTC n=126
O	1,7	0,0
C	10,3	15,1
FS	86,0	78,6
M	1,2	5,6
U	0,7	0,8

TABLEAU 3.3b. APTITUDES À LA LECTURE

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=541	JTC n=135
O	2,6	1,5
C	11,8	9,6
FS	79,9	83,7
M	2,8	5,2
U	3,0	0,0

TABLEAU 3.3d. ENTRÉE DES RÉPONSES

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=569	JTC n=137
O	1,2	0,0
C	8,3	5,8
FS	88,4	82,5
M	1,1	8,8
U	1,1	2,9

TABLEAU 3.3e. CAS PROBLÈMES

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=58	JTC n=35
O	25,9	8,6
C	34,5	28,6
FS	39,7	60,0
M	0,0	2,9
U	0,0	0,0

TABLEAU 3.3f. MAÎTRISE DES CONCEPTS

CONTRÔLE SYSTÉMATIQUE	% DES INTERVIEWEURS NOTÉS	
	TTC n=504	JTC n=122
O	6,2	0,8
C	5,0	13,1
FS	85,9	84,4
M	2,2	1,6
U	0,8	0,0

Nous présentons au tableau 3.4 les moyennes des notes accordées à la séance de contrôle (et les valeurs des écarts-types) pour chaque catégorie de contrôles. L'utilisation de SAS Proc TTEST pour évaluer la signification statistique des différences entre les moyennes selon le centre téléphonique laisse conclure à l'existence de différences pour certaines catégories de contrôles. Pour les contrôles initiaux, on observe des différences statistiquement significatives entre les notes moyennes données aux deux centres pour les cas problèmes. Dans le cas des contrôles systématiques, on observe une différence statistiquement significative pour la note accordée pour les aptitudes à la lecture, l'entrée des réponses et les moyennes pour la séance. Ce tableau montre également qu'après le contrôle initial et la rétroaction, les intervieweurs qui passent sous le régime du contrôle systématique fournissent un travail bien supérieur, ce qui porte à conclure à l'efficacité d'ensemble des contrôles réalisés aux deux centres.

TABLEAU 3.4. VALEUR MOYENNE DES NOTES DE CONTRÔLES PAR CATÉGORIE

CATÉGORIE	CONTRÔLE INITIAL		CONTRÔLE SYSTÉMATIQUE	
	TTC	JTC	TTC	JTC
VOIX/FAÇON DE S'EXPRIMER	2,02 (0,25)	1,96 (0,25)	3,17 (0,51)	3,12 (0,46)
APTITUDES À LA LECTURE	1,88 (0,48)	1,88 (0,40)	3,08 (0,60)	3,07 (0,45)
APTITUDES À APPROFONDIR	1,87 (0,40)	1,86 (0,34)	3,11 (0,45)	3,08 (0,48)
ENTRÉE DES RÉPONSES	1,94 (0,38)	1,84 (0,37)	3,08 (0,42)	2,91 (0,51)
CAS PROBLÈMES	2,33 (0,49)	1,93 (0,26)	3,86 (0,80)	3,43 (0,70)
MAÎTRISE DES CONCEPTS	1,91 (0,38)	1,79 (0,41)	3,13 (0,58)	3,13 (0,41)
MOYENNE DE LA SÉANCE	—	—	3,13 (0,40)	3,07 (0,25)

3.3 Résultats des contrôles selon l'évaluateur

TABLEAU 3.6a. NOTES MOYENNES ACCORDÉES AUX CONTRÔLES SYSTÉMATIQUES PAR ÉVALUATEUR ET PAR CATÉGORIE - TTC

ÉVALUA-TEUR	VOIX/FAÇON DE S'EXPRIMER	APTITUDES À LA LECTURE	APTITUDES À APPROFONDIR	ENTRÉE DES RÉPONSES	CAS PROBLÈMES	MAÎTRISE DES CONCEPTS	NOTE GLOBALE
W02	3,02	2,96	3,00	2,95	4,50	3,04	3,00
W09	2,96	2,96	3,00	2,96	3,67	2,96	2,99
W10	3,16	3,15	3,06	3,07	3,29	3,04	3,09
W12	2,94	2,88	2,86	2,97	5,00	2,90	2,93
W13	3,59	3,33	3,52	3,49	4,50	3,79	3,54
W14	3,10	3,19	2,96	2,95	3,57	2,90	3,02
W16	3,02	2,86	3,06	2,91	3,50	2,97	2,97
W20	3,12	3,17	3,47	3,06	4,33	3,29	3,22
W21	3,65	3,55	4,00	3,43	4,75	3,48	3,60
W24	3,00	3,00	3,00	3,00	—	3,00	3,00
W26	3,12	2,92	3,03	3,00	3,50	3,00	3,01
W27	3,03	2,90	3,03	3,00	4,20	3,08	3,02
W29	3,15	3,10	3,03	3,05	4,00	3,07	3,08

La question de l'uniformité des évaluations réalisées par les moniteurs/surveillants est une préoccupation constante dans la conduite des ITAO. Comme les résultats des contrôles servent à l'évaluation du rendement des intervieweurs, la fiabilité de ces contrôles prend un sens particulier au Bureau. Nous présentons aux tableaux 3.6a et 3.6b les valeurs moyennes des notes accordées aux contrôles systématiques par chacun des évaluateurs au cours de la période à l'étude dans les deux centres téléphoniques. L'analyse à l'aide de SAS Proc GLM donne à conclure qu'à Jeffersonville, il existe une différence significative entre les notes accordées pour les aptitudes à

approfondir, l'entrée des réponses et la maîtrise des concepts d'un évaluateur à l'autre. Dans l'ensemble, les résultats statistiques laissent constater une plus grande variance dans les notes accordées à Tucson que dans celles accordées à Jeffersonville.

TABLEAU 3.6b. NOTES MOYENNES ACCORDÉES AUX CONTRÔLES SYSTÉMATIQUES PAR ÉVALUATEUR ET PAR CATÉGORIE - JTC

ÉVALUA- TEUR	VOIX/FAÇON DE S'EXPRIMER	APTITUDES À LA LECTURE	APTITUDES À APPROFONDIR	ENTRÉE DES RÉPONSES	CAS PROBLÈMES	MAÎTRISE DES CONCEPTS	NOTE GLOBALE
SW12	3,03	3,04	3,04	2,94	3,75	3,07	3,02
SW15	3,00	2,50	1,00	1,50	—	3,00	2,44
SW17	3,32	3,12	3,27	2,71	3,46	3,18	3,11
SW18	3,17	3,17	3,33	3,00	4,50	3,00	3,21
SW19	3,71	3,57	3,25	3,57	3,67	3,86	3,63
SW20	3,00	3,00	3,00	3,00	3,08	3,00	3,02
SW21	3,00	3,00	3,00	3,00	—	4,00	3,17

4. CONCLUSION

Les contrôles réalisés aux centres ITAO du *Census Bureau* constituent des outils utiles de gestion des enquêtes. Ils permettent d'évaluer le rendement des intervieweurs ainsi que celui des instruments et des méthodes d'enquêtes. Ces contrôles, utilisés aux fins de la rétroaction, permettent au surveillant ou au moniteur d'observer discrètement le déroulement des interviews et de fournir une rétroaction immédiate et précise. Ils permettent également de mesurer précisément l'utilité des instruments d'enquêtes en conditions réelles, et d'apporter des améliorations aux questionnaires actuels et futurs. Les informations obtenues grâce aux contrôles servent à modifier la formation des intervieweurs ainsi qu'à réduire les erreurs non dues à l'échantillonnage dans les enquêtes. Nous avons relevé l'existence de certaines différences entre les centres ITAO quant à la façon de réaliser les contrôles et d'utiliser les données de ces contrôles. Nous suggérons d'examiner plus à fond ces différences et de faire en sorte que tous les centres puissent bénéficier des résultats obtenus.

5. BIBLIOGRAPHIE

Biemer, Paul (1996). « Evaluator Error in the Assessment of Interviewer Performance », *Bureau of the Census*, 10 juillet.

Cannell, C.F. et Oksenberg, L. (1988). « Observation of Behavior in Telephone Interviews », in R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls et J. Waksberg (éds), *TELEPHONE SURVEY METHODS*, p. 475-495, New York: Wiley.

Couper, Mick, Groves, Robert M. et Smith, Timothy K. (1990). « Developing Systematic Procedures for ITAO Monitoring », Exposé présenté à la Conférence internationale sur les erreurs de mesures, Tucson, AZ. Novembre.

Couper, M.P., Holland, L. et Groves, R.M. (1992). « Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility », *JOURNAL OF OFFICIAL STATISTICS*, Vol. 8, No. 1, 63-76.

SESSION 5

NOUVELLES MÉTHODES DE COLLECTE ET SAISIE

EXPÉRIENCES DE STATISTIQUE CANADA AVEC LA SAISIE AUTOMATISÉE DE DONNÉES

Suzanne M. Vézina¹

RÉSUMÉ

Au cours des dernières années, Statistique Canada a procédé à des travaux de recherches dans le domaine de la saisie automatisée de données. Environ deux années ont été consacrées à la recherche de la technologie et la dernière année a été consacrée au développement et à la mise en œuvre de systèmes de saisie automatisée de données pour diverses enquêtes du bureau.

Le but de la communication est de partager certains des résultats obtenus. Ces résultats font ressortir l'importance de la conception des formulaires et des questionnaires pour assurer des données de qualité. Le formulaire est l'élément essentiel de tout système de saisie de données. Un formulaire bien conçu facilite la cueillette et la saisie de données et il doit aussi être facile à remplir. La saisie automatisée de données est optimisée sur le plan qualité lorsque les secteurs réservés aux réponses sont clairement définis. Le procédé automatisé utilisant une image du formulaire comme agent de saisie, le papier et l'encre employés doivent produire une image de bonne qualité.

Des formulaires conçus afin de favoriser la cueillette et la saisie automatisée minimisent les erreurs introduites lors de ces deux procédés. La saisie automatisée permet finalement d'accélérer le procédé tout en produisant des données ayant les normes de qualité attendues.

MOTS CLÉS: saisie automatisée de données; conception de formulaires; image.

1. INTRODUCTION

La saisie dite traditionnelle est celle où les réponses sont entièrement entrées ou tapées au clavier par un opérateur, à partir d'un formulaire. Ce procédé manuel est long, répétitif et monotone. Il favorise par conséquent l'introduction d'erreurs. Des procédés permettent le contrôle de la qualité permettent la diminution et la presque élimination des erreurs de frappe. Le contrôle de qualité est une composante additionnelle liée à la saisie de données. À cette composante se rattache également l'élément de temps et de coût.

L'intention est de minimiser les erreurs introduites dans les données lors de la cueillette et lors de la saisie. La saisie automatisée de données à partir de formulaires conçus selon des normes pré-établies nous offre une avenue intéressante.

2. SAISIE AUTOMATISÉE DE DONNÉES

La saisie automatisée de données en est une où l'interprétation des réponses et par ce fait l'emmagasinage sur ordinateur est majoritairement effectuée par un ordinateur et complétée ou corrigée par un opérateur. Tout le processus est effectué à partir d'images de formulaires.

¹Suzanne M. Vézina, Statistique Canada, Ottawa, Ont, K1A 0T6.

Les fonctions d'un système automatisé de données sont:

- ▶ La numérisation ou l'importation introduction des formulaires dans le lecteur optique qui lit (numérise) et enregistre les images dans un format exploitable par un ordinateur. Des images déjà créés peuvent aussi être importées dans un système de saisie automatisée de données.
- ▶ L'identification du formulaire identification automatique des formulaires (pages) soumis à la saisie automatisée.
- ▶ L'exception repérage automatique des pages qui ne font pas partie du formulaire et des pages qui n'ont pas été bien identifiées par le lecteur optique.
- ▶ L'extraction reconnaissance automatique des caractères imprimés à la machine, des caractères imprimés à la main, des codes à barres et des marques.
- ▶ La vérification vérification et correction manuelles des données suspectes et des données qui ne satisfont pas aux règles de validation. Ces données ainsi que l'image du formulaire sont affichées sur les postes de travail des vérificateurs. Les vérificateurs complètent les données à l'aide de ces images.

3. FACTEURS QUI FAVORISENT LA CUEILLETTE ET LA SAISIE DE DONNÉES

Le formulaire est l'élément clé de tout système de saisie de données. Un formulaire conçu en considérant les facteurs qui influencent la cueillette et par conséquent la saisie, diminue les erreurs introduites dans les données lors de cette cueillette et aussi lors de la saisie subséquente des données. Ces affirmations sont vraies que la saisie soit effectuée de façon manuelle ou automatique.

La saisie manuelle de données étant effectuée par une personne, cette dernière tente de déchiffrer les données à saisir lorsque les réponses sont difficiles à lire. Les réponses peuvent être soit tout simplement difficiles à lire, soit écrites totalement ou partiellement à l'extérieur des régions ou zones réservées aux réponses, ou même soit en contravention à ce qui est attendu comme réponse à une question. L'oeil humain pouvant s'adapter, l'opérateur de saisie manuelle accomplit son travail du mieux qu'il peut.

Une conception de formulaire selon des paramètres et des normes rigoureuses contribue à diminuer les erreurs introduites lors de la cueillette. À partir d'un tel cadre, la saisie automatisée devient alors une alternative réalisable. La conception des formulaires de façon à favoriser la cueillette et par conséquent la saisie automatisée est essentielle au succès du processus de saisie automatisée. Cette saisie étant majoritairement effectuée par une machine, le procédé s'effectue de façon accélérée et offre une constance dans les résultats. La possibilité de définir des niveaux de confiance pour les résultats de la reconnaissance contribue à la diminution des erreurs introduites lors de la saisie. Dans un tel environnement les opérateurs n'interviennent que lorsque la machine questionne les résultats de la reconnaissance, ou lorsque les résultats contreviennent aux règles de validation définies dans l'application. Des formulaires conçus selon des normes qui favorisent la cueillette, des images de bonne qualité, un système de saisie automatisée reconnu et performant et finalement des règles de validations qui sont appliquées aux données extraites sont donc tous des facteurs qui diminuent les erreurs introduites lors de la cueillette et la saisie de données.

Les facteurs reliées à la conception du questionnaire qui favorisent la cueillette et par conséquent la saisie automatisée sont les caractéristiques du papier, les couleurs d'encre utilisées pour imprimer le formulaire, les identificateurs de page, les points d'ancrage, la définition des zones de données à saisir et finalement l'impression des formulaires.

4. CARACTÉRISTIQUES DU PAPIER

Tous les lecteurs optiques, les logiciels de balayage et les logiciels de saisie automatisée de données présentent des limites en ce qui concerne le papier. La texture, la teneur en fibres, la qualité de la surface, l'épaisseur ou le poids, le pouvoir réfléchissant, l'opacité, la couleur et les dimensions sont tous des facteurs importants.

- La texture, la teneur en fibres et la qualité de la surface influencent la clarté et précision de l'image produite.
- L'épaisseur ou poids ont un impact direct sur le fonctionnement du lecteur optique. Une numérisation continue, sans heurt et sans interruption est souhaitable. Si trop épais ou trop mince, la numérisation sera continuellement interrompue ou même impossible. Lorsque le papier est trop épais le mécanisme du lecteur optique ne pourra pas l'accepter. Lorsque le papier est trop mince et que le lecteur réussit la numérisation, les images sont souvent floues à cause de l'ondulation créée lorsque le formulaire est soumis au mécanisme du lecteur optique.
- Un papier non lustré doit être utilisé. Le papier lustré cause de la réflexion lors de la numérisation ce qui embrouille des parties de l'image.
- Le papier utilisé doit être opaque afin d'éviter que les écritures d'un côté d'une page n'apparaissent sur l'image de l'autre côté de cette page.
- Le papier devrait être de couleur blanche ou pastel, avec suffisamment de réflectivité, afin que les images apparaissent blanches lors du visionnage à l'écran.

Le papier ordinaire produit à partir de fibres non recyclées ainsi que la plupart des papiers fabriqués à partir de fibres recyclées sont ordinairement utilisables pour l'imagerie.

5. COULEURS D'ENCRE

L'utilisation des couleurs ainsi que l'utilisation de régions sans couleur contribuent à la bonne apparence des formulaires et par conséquent des images prises à partir de ces formulaires. Les couleurs guident les répondants, diminuant ainsi les ambiguïtés et par le fait même l'introduction d'erreurs au niveau de la cueillette. Les couleurs d'encre utilisées sur les formulaires peuvent être non discernées ou discernées lors de la numérisation des formulaires. Les couleurs d'encre non discernées utilisées autour de régions de réponses facilitent la saisie automatisée de données.

- Les couleurs d'encre non discernées sont de teintes subtiles. Elles sont utilisées afin de guider le répondant sur le formulaire et sont invisibles au lecteur optique. Elles ne sont pas perçues par le lecteur optique parce que leur degré de réflectivité est élevé. L'utilisation de ces couleurs d'encre est appropriée pour encadrer les zones de réponses ainsi que pour toute autre information non essentielle comme les arrières-plans et les graphiques.

L'utilisation de couleurs d'encre non discernées aux endroits appropriés améliore la reconnaissance des données, réduit la taille des images ce qui diminue les besoins d'espace d'emmagasinage, accélère le traitement des images et produit finalement des images plus claires lors du visionnage à l'écran.

- Les couleurs d'encre discernées sont de teintes foncées. Les couleurs d'encres discernées sont utilisées:
 - pour imprimer les identificateurs de pages et les points d'ancrage.

- pour toute information qui doit être extraite par le système de reconnaissance.
- pour toute information qui doit être vue lorsque l'image est visionnée à l'écran, par exemple les questions.

Le noir est la couleur la plus utilisée pour l'impression d'information qui doit être discernée lors de la numérisation de formulaires. Il est important que l'information imprimée avec des couleurs d'encre discernées n'obstruent pas les zones de réponses.

Il faut aussi définir des espaces sans couleur lors de la conception d'un formulaire. Les zones d'extraction et leur environnement immédiat ne doivent pas être obstrués ni embrouillés par des couleurs partiellement ou entièrement discernées.

Un autre point à considérer est la région de pliure d'un formulaire. Lorsqu'un formulaire a été plié, l'image produite lors de la numérisation sera grisâtre dans la région du pli. Il est donc important de ne pas concevoir de zone de réponses dans cette région d'un formulaire.

6. IDENTIFICATEURS DE PAGES

Les identificateurs de pages permettent la reconnaissance automatique des formulaires soumis au système de saisie automatisée de données. Le système examine l'image et détermine quel formulaire et quelle page à l'intérieur du formulaire sont soumis au traitement. Les identificateurs de pages permettent la détection de pages qui ne font pas partie du formulaire traité, lorsque de telles pages ont été introduites involontairement.

Un identificateur de page doit être unique à l'intérieur d'un formulaire de plusieurs pages. Un identificateur de page doit être défini pour chaque côté de chaque page d'un formulaire. Les identificateurs de pages doivent être imprimés avec une couleur d'encre discernable par le lecteur optique. Voici quelques exemples d'identificateur de page: EERH, E1, E2.

7. POINTS D'ANCRAGE

Les points d'ancrage permettent au système de bien aligner les images et d'éliminer la mise en travers au cours du balayage. L'alignement des images est essentiel pour obtenir de bons résultats lors de l'extraction ou reconnaissance des données.

Des points d'ancrages doivent être déterminés sur chaque côté de chaque page d'un formulaire. Ils doivent être imprimés avec une couleur d'encre discernable par le lecteur optique. Des points d'ancrage doivent être imprimés à divers endroits de la page. Le système de saisie automatisée de données utilisé détermine le genre et les emplacements possibles des points d'ancrage sur une page. Des exemples de points d'ancrage utilisés par plusieurs systèmes de saisie automatisée de données sont des petits triangles noirs ou des cercles noirs imprimés à chaque coin d'une page.

8. DÉFINITION DES ZONES DE DONNÉES À SAISIR

Une zone réservée aux données est conçue en fonction du type de données que l'on cherche à obtenir. Le tableau suivant énumère le type de données qui peuvent être reconnues par un système de saisie automatisée de données. Il spécifie également en français et en anglais, le genre de technologie utilisée pour la reconnaissance de chacune.

Quelques notions au sujet de la conception d'un formulaire en fonction du type de données recherchées suivent le tableau.

TYPES DE DONNÉES	TECHNOLOGIE - Français	TECHNOLOGIE - Anglais
Codes à barres	RCB Reconnaissance de Codes à Barres	BCR Bar Code Recognition
Caractères imprimés avec un jeu de caractères à l'encre magnétique	RCEM Reconnaissance de Caractères à Encre Magnétique	MICR Magnetic Ink Recognition
Marques	ROM Reconnaissance Optique de Marques	OMR Optical Mark Recognition
Caractères imprimés à la machine	ROC Reconnaissance Optique de Caractères	OCR Optical Character Recognition
Caractères imprimés à la main	RIC Reconnaissance Intelligente de Caractères	ICR Intelligent Character Recognition

8.1 Codes à Barres

Les codes à barres correspondent à des données alphanumériques et font appel à une série de lignes verticales de différentes largeurs qui sont séparées par des espaces:



Les codes à barres doivent être imprimés avec une couleur d'encre discernable par le lecteur optique. Une région blanche doit exister tout autour des codes à barres, c'est-à-dire qu'aucune autre information imprimée à l'encre discernée ne doit obstruer les codes eux-mêmes.

8.2 Caractères imprimés avec un jeu de caractères à l'encre magnétique

Ces caractères sont formés avec un jeu de caractères hautement spécialisés et ils sont imprimés à l'encre magnétique:



Ce genre de caractères doit être imprimé avec une couleur d'encre discernable par le lecteur optique. Une région blanche doit exister tout autour des caractères, c'est-à-dire qu'aucune autre information imprimée à l'encre discernée ne doit obstruer les caractères eux-mêmes.

8.3 Marques

Les zones ou champs à marques sont des secteurs délimités dans lesquels le répondant indique un choix en faisant une marque:

- Option A
- Option B
- Option C

- Oui
- Non

Les zones à marques doivent être délimitées soit par une boîte, un cercle ou une ellipse. Selon le système de saisie automatisée de données, ces derniers seront imprimés soit avec de l'encre à couleur discernée soit avec de l'encre à couleur non discernée. La dimension idéale pour une zone à marques est 1/8 de pouce carré. Elle peut aussi être plus grande. Il est important de distancer les zones les unes des autres afin d'éviter qu'un choix indiqué par le répondant ne déborde pas dans une autre zone de réponse.

8.4 Caractères imprimés à la machine

Les caractères imprimés à la machine comprennent tous les jeux de caractères standards ainsi que les caractères typographiques et les caractères produits par des imprimantes au laser et des machines à écrire:

QUEST1234567890

Une région blanche qui mesure au moins 1/4 de pouce doit exister tout autour des caractères, c'est-à-dire qu'aucune autre information imprimée à l'encre discernée ne doit obstruer les caractères eux-mêmes.

L'espacement entre les caractères, l'espacement entre plusieurs lignes de caractères ainsi que la densité des caractères influencent la performance de la reconnaissance. En fait, si les caractères se touchent ou s'ils ne sont pas suffisamment à l'écart de toute autre écriture ou diagramme, la reconnaissance sera difficile.

8.5 Caractères imprimés à la main

Comme le nom l'indique, les caractères imprimés à la main sont des caractères qui ont été imprimés par une personne.

SUZANNE

1 2 3 4 5 6

S U Z A N N E

Les caractères imprimés à la main sont ceux qui sont le plus difficile à reconnaître automatiquement parce qu'ils sont très variés et inconsistants des points de vue de styles et de formes. Les caractères imprimés à la main sont plus clairs et par conséquent plus faciles à lire lorsque la conception des zones pour réponses imprimées à la main sont encadrées. Plus les cadres d'un formulaire sont précis, plus les répondants sont impulsivement portés à écrire au bon endroit, de façon claire et constante. Le moyen le plus propice pour recevoir des données imprimées à la main est un ensemble de boîtes. Les boîtes incitent les répondants à imprimer les caractères en majuscules au lieu d'écrire avec le style manuscrit ou cursif.

- Prévoir suffisamment de boîtes pour recevoir le nombre maximum de caractères pour une réponse.
- La grandeur minimum des boîtes qui permet au répondant de bien former les caractères est 3/16 de pouce de largeur par 1/4 de pouce de hauteur.
- Les boîtes peuvent être côte à côte, c'est-à-dire sans espace entre elles. Dans ce cas, la ligne médiane qui sépare deux boîtes est unique et donc partagée par celles-ci.
- Les boîtes peuvent être juxtaposées, c'est-à-dire séparées légèrement les unes des autres.
- Lorsque plusieurs rangées de boîtes se superposent, ces rangées doivent être séparées les unes des autres par un minimum de 1/8 de pouce.
- Aucune information ne doit obstruer l'espace réservé pour les caractères.

9. IMPRESSION DES FORMULAIRES

Comme l'image est le médium à partir duquel la saisie automatisée de données est effectuée, la qualité d'impression de celle-ci influence directement la performance du système.

Le positionnement précis des identificateurs de formulaires, des points d'ancrage ainsi que des zones de réponses, ceci *sur toutes les copies d'un même formulaire* sont les éléments clés qui permettront une saisie efficace et idéalement exempte d'erreurs. Plus les différentes zones sont clairement et précisément délimitées d'un formulaire à l'autre (ceci pour un même formulaire), plus la saisie sera performante.

Les lignes et boîtes qui sont imprimées avec de l'encre à couleur discernée et qui sont situées autour des caractères à extraire doivent être continues et d'épaisseur constante. Plusieurs logiciels de saisie automatisée utilisent une technique qui consiste à enlever les boîtes avant de procéder à l'extraction. Lorsque les boîtes sont enlevées, elles risquent moins d'interférer avec les caractères. Lorsque les lignes et les boîtes ne sont pas imprimées de façon continue et avec une épaisseur constante, ces logiciels ne les reconnaissent pas efficacement. Ils ne les enlèvent donc pas parce que s'ils le faisaient, ils risqueraient d'enlever par le fait même des caractères à extraire. De telles situations n'existent pas lorsque les lignes et les boîtes sont imprimées avec de l'encre à couleur non discernée.

10. CONCLUSION

La saisie automatisée de données est une technologie viable qui permet de réaliser des bénéfices. La conception de formulaires selon les normes citées dans cet article est une partie essentielle et intégrante de l'exploitation d'un tel système. Des répondants motivés écrivent plus clairement, avec plus de régularité et de constance. Moins d'erreurs sont alors introduites lors de la cueillette. L'automatisation de la saisie de données devient possible avec ces formulaires et cette automatisation résulte en une diminution du nombre d'erreurs par rapport à une saisie qui est entièrement effectuée de façon manuelle.

Les bénéfices sont réalisés entre autres au niveau du temps de saisie qui est plus court, au niveau de la qualité des données qui est améliorée, et enfin au niveau du coût du procédé qui est moindre que si la saisie est effectuée

manuellement:

- ▶ La saisie de données effectuée par ordinateur accélère de beaucoup le temps consacré au procédé. La vitesse à laquelle les ordinateurs fonctionnent ne cesse d'augmenter. Il est de plus en plus avantageux de les utiliser pour tout traitement de données, dont la saisie de données.
- ▶ Le fait que les agents de saisie n'interviennent que dans les cas où le système ne peut extraire les données avec certitude contribue également à la diminution du temps requis pour compléter le procédé. La diminution du nombre d'intervention manuelles résulte en une diminution du nombre d'erreurs introduites lorsque le procédé est complètement manuel.
- ▶ La saisie automatisée à partir d'un formulaire conçu selon les normes citées dans cet article et l'application de règles de validations produisent des données qui contiennent moins de 1 % d'erreurs. Ce pourcentage est basé sur l'hypothèse que les agents de saisie n'introduisent pas d'erreurs à l'étape de vérification. La vérification étant cependant un procédé manuel, les agents de saisie introduisent parfois des erreurs. Les validations qui sont appliquées aux données extraites permettent de minimiser le nombre d'erreurs introduites lors du procédé de vérification. Des techniques de contrôle de la qualité sont également incluses dans plusieurs logiciels de saisie automatisée de données. Ces techniques permettent de détecter et de corriger directement les erreurs de saisie de données.
- ▶ L'économie de temps se révèle également être une économie d'argent. L'investissement initial en logiciel et en matériel informatique est récupéré selon l'utilisation qui est faite du système. Selon le système de saisie automatisée utilisé, au moins 40 % des agents de saisie peuvent être affectés à des tâches différentes. Ce chiffre peut même passer à 60 % avec l'utilisation de systèmes de saisie automatisée de haute gamme.

ERREUR NON DUE À L'ÉCHANTILLONNAGE : LA DÉCLARATION ÉLECTRONIQUE PEUT-ELLE AIDER?

Laurie Hill¹

RÉSUMÉ

La déclaration électronique des données est encore relativement nouvelle à Statistique Canada, mais les résultats préliminaires d'un certain nombre d'essais laissent supposer que pour les enquêtes auprès des entreprises, elle pourrait contribuer à augmenter l'actualité des données, à réduire les coûts et à améliorer la qualité des données, du fait de la réduction du nombre d'erreurs non dues à l'échantillonnage.

MOTS CLÉS: non-réponse; erreur; électronique; déclaration.

1. INTRODUCTION

1.1 En quoi consiste la déclaration électronique des données?

Pendant des années, la collecte des données à Statistique Canada, dans le cadre des enquêtes- entreprises, a principalement pris la forme de questionnaires sur papier, envoyés par la poste aux répondants et renvoyés par la poste, une fois remplis, par ceux-ci. Plus récemment, on a mis en oeuvre un certain nombre d'approches de remplacement: le suivi par téléphone au sujet des questionnaires envoyés par la poste, la collecte des données par téléphone, des intervieweurs notant sur papier les renseignements fournis, en vue de la saisie subséquente des données, les interviews en personne par les employés de Statistique Canada et les interviews téléphoniques assistées par ordinateur (ITAO). On a aussi fait l'expérience de la collecte des données au moyen du téléphone à boutons, de l'interview en personne assistée par ordinateur et d'autres techniques innovatrices. Nous utilisons maintenant ce que nous appelons généralement une méthode "mixte" de collecte des données. Nous continuons d'utiliser un certain nombre des techniques énoncées précédemment, que nous combinons pour obtenir la méthode la plus efficace de collecte des données requises, au moment opportun, tout en réduisant le fardeau de réponse.

La déclaration électronique des données est la plus nouvelle des techniques dont nous faisons l'expérience, dans un effort continu en vue de réduire le fardeau de déclaration qui semble être imposé aux répondants, tout en améliorant la rapidité et l'efficacité de nos activités de collecte.

Nous avons mis à l'essai un certain nombre de méthodes de communication électronique des questionnaires et de récupération des données par la même voie. La méthode la plus utilisée jusqu'à présent a consisté à présenter les questions d'enquête et à préciser la forme et le contenu des réponses sur des disquettes envoyées par la poste aux répondants. Les répondants utilisent leur ordinateur pour remplir les questionnaires électroniques et renvoient généralement les données à Statistique Canada sur disquette. Nous mettons actuellement au point d'autres options, en vue d'offrir aux répondants davantage de choix quant aux méthodes de transmission, y compris l'impression des données et leur envoi par télécopieur, l'envoi de fichiers par courrier électronique, le transfert de modem à modem et (bientôt) le transfert protégé au moyen d'Internet. Nous envisageons en outre la possibilité de tirer des données des dossiers administratifs sur support électronique des répondants, en obtenant au préalable la permission de ces derniers, ou encore d'obtenir des copies des transactions électroniques à partir des "réseaux à valeur ajoutée", qui servent à la transmission de ces transactions entre les "partenaires commerciaux".

¹ Laurie Hill, Directeur adjoint, Division des enquêtes-entreprises, Statistique Canada, Parc Tunney, Ottawa, Canada, K1A 0T6.

Nos objectifs sont toujours les mêmes : réduire le fardeau réel ou perçu de réponse pour les entreprises répondantes et, parallèlement, obtenir des réponses de meilleure qualité à nos enquêtes. La déclaration électronique fait partie des options que nous offrons aux répondants, mais nous n'insistons pas pour qu'ils l'utilisent.

1.2 TRAITS COMMUNS DE CES TECHNIQUES

- Elles sont toutes conçues pour faciliter le travail du répondant.
- Nous précisons les données que nous voulons recevoir sur support électronique et nous indiquons aussi sous quelle forme nous les voulons.
- Les vérifications sont intégrées aux applications, afin de pouvoir être effectuées au moment de l'entrée des données par les répondants dans le programme, et non pas à la réception de celles-ci à Statistique Canada.
- Jusqu'à maintenant, tous les efforts en matière de déclaration électronique ont été axés sur les enquêtes, c'est-à-dire que l'on n'a pas élaboré de systèmes généralisés. Les applications sont conçues pour une seule enquête, et la plupart des travaux de conception doivent être repris si l'on veut appliquer la technique à une autre enquête.

1.3 AVANTAGES POSSIBLES POUR STATISTIQUE CANADA

- Actualité améliorée, pourvu qu'on y mette les efforts nécessaires!
- Coûts d'enquête plus faibles, si les applications sont bien conçues, et seulement après une période d'introduction.
- Qualité améliorée des données, si les vérifications appropriées sont intégrées à l'application.

1.4 RÉSULTATS

En ce qui a trait à la question de l'actualité, les avis sont encore partagés. Il ne fait aucun doute que certains des rapports nous sont revenus plus rapidement sur support électronique que lorsqu'ils étaient remplis par les mêmes répondants au moyen de questionnaires sur papier. Toutefois, il existe aussi des cas où les répondants ont pris plus de temps pour remplir le questionnaire électronique que celui sur papier. De façon générale, nous pouvons dire qu'il faut continuer d'assurer un suivi relativement aux techniques de déclaration électronique. Au cours de la période d'introduction, il est essentiel que les répondants disposent d'une "ligne téléphonique d'aide", étant donné que peu importe le niveau de détail des instructions, les répondants continueront de se heurter à des problèmes. Il faudra encore assurer beaucoup de soutien, et si les répondants n'ont pas "quelqu'un vers qui se tourner", il se peut qu'ils abandonnent en cours de route. Si l'on veut que la déclaration électronique fonctionne, on doit être prêt à investir du temps et des efforts, non seulement pour élaborer une bonne application, mais aussi pour aider les répondants à se familiariser avec elle et à l'utiliser dans leur environnement. Les retombées du point de vue de l'amélioration de l'actualité devraient venir avec le temps.

En ce qui a trait aux coûts, nous savons que des économies peuvent être réalisées. Après tout, la saisie des données est effectuée par les répondants, ce qui fait que les économies devraient être considérables, pour cette fonction à tout le moins. En outre, pour la plupart des "bonnes" applications de déclaration électronique, le fait que le répondant effectue lui-même la vérification des données devrait aussi entraîner une réduction des coûts. Cela est vrai si l'application est bien conçue, si les vérifications y sont intégrées, et si les répondants comprennent ce que l'on attend d'eux. En résumé, il fait très peu de doute que la déclaration électronique des données entraînera une réduction des coûts, mais seulement si l'on investit

suffisamment de ressources dans la conception d'une application appropriée et si l'on planifie bien l'entrée en vigueur des nouvelles méthodes.

Toutefois, la question véritable qui se pose dans le cadre du présent symposium est la suivante : "La déclaration électronique des données entraîne-t-elle une amélioration de la qualité de celles-ci?". Le cas échéant, les améliorations prendront sans aucun doute la forme d'une diminution de l'erreur non due à l'échantillonnage. Il n'existe pas de différence du point de vue de l'échantillonnage lorsque l'on utilise des techniques de déclaration électronique, non plus que du point de vue de la base de sondage. Les améliorations découleront, le cas échéant, de la suppression des erreurs introduites par l'intervieweur, ainsi que d'une réduction des erreurs d'interprétation de la part des répondants, si la conception de l'application répond aux conditions énoncées précédemment.

Il est trop tôt pour dire avec certitude que la qualité des données s'est améliorée grâce aux techniques de déclaration électronique. Les applications utilisées jusqu'à maintenant n'ont pas été conçues dans une optique d'évaluation, ce qui fait qu'il est difficile de tirer des conclusions. Nous aimerions disposer de statistiques comparables sur des rejets à la vérification touchant des enregistrements tirés d'échantillons de répondants, avant et après l'application de techniques de déclaration électronique. Malheureusement, dans les applications utilisées jusqu'à maintenant, on n'a pas conservé de tels enregistrements. Toutefois, les résultats de l'application du PERQS (Système de gestion de questionnaires électroniques personnalisés) à l'Enquête annuelle sur les magasins à succursales et les grands magasins sont prometteurs. Des techniques de déclaration électronique ont été mises à l'essai dans le cadre de cette enquête auprès d'un échantillon de 25 répondants importants, pour l'enquête de 1994, et l'application a été élargie à 280 entreprises répondantes pour l'enquête de 1995. Ces 280 répondants représentent plus de 50 % des points de vente au détail visés par l'enquête. Les superviseurs concernés de notre Division des opérations et de l'intégration sont certains que la qualité s'est améliorée. Ils disent que la **proportion de fichiers épurés est en hausse**, étant passée de 45 %, au moyen des techniques utilisées précédemment (un mélange de collecte par la poste et d'ITAO), à près de 75 % au moyen de la collecte par déclaration électronique sur disquette. Ils soulignent en outre qu'une grande partie des autres 25 % d'enregistrements qui échouent à la vérification viennent de répondants qui passent outre aux vérifications de l'application, mais ne fournissent pas de raison ni d'explication à ce sujet. Des observations comme celles-ci sont réconfortantes, mais il serait encore plus rassurant de disposer de preuves statistiques étayées selon lesquelles la qualité s'est définitivement améliorée. Des méthodes visant à évaluer les changements qui se sont produits du point de vue de la qualité devraient de toute évidence faire partie de la prochaine application d'essai de la déclaration électronique des données.

1.5 Orientations pour l'avenir

Statistique Canada a décidé récemment de mettre en oeuvre un nouveau programme de relations avec les répondants pour les grandes entreprises. Ces grandes entreprises sont visées par nombre de nos enquêtes, et leurs activités sont tellement vastes qu'il est essentiel qu'elles répondent, et qu'elles répondent correctement. Ces entreprises sont perçues comme des "fournisseurs clés de données", et Statistique Canada est à mettre en oeuvre un programme de gestion des fournisseurs clés pour mieux se concentrer sur elles. Chacune des entreprises choisies se verra affecter un gestionnaire de fournisseur clé (GFC), c'est-à-dire un responsable de Statistique Canada qui s'occupera en priorité d'un nombre restreint de ces répondants importants. Le GFC sera chargé de communiquer aux entreprises un aperçu global de nos exigences en matière de données, de coordonner la collecte des données pour toutes les enquêtes-entreprises parrainées par le Bureau ainsi que de négocier les modalités de déclaration des données avec les entreprises, pour réduire le fardeau de réponse. On s'attend à ce que les GFC fassent notamment la promotion de la déclaration électronique de données, en vue de réduire le fardeau de réponse. Le GFC sera en outre responsable de la coordination de la comparaison des données avec Statistique Canada pour tous les rapports reçus d'entreprises dont il est chargé. Ainsi, nous nous attendons à améliorer la cohérence, la portée et la qualité des réponses que nous recevons de ces grandes entreprises.

Dans le cas des petites et moyennes entreprises, nous n'avons pas de programme similaire pour le moment. Toutefois, certaines suggestions ont été faites qui, si elles sont mises en oeuvre, devraient entraîner des améliorations. Nous pourrions commencer à élaborer un ensemble plus cohérent d'exigences en ce qui a

trait aux données, sur une base mensuelle, trimestrielle et annuelle. Nous pourrions par la suite relier le plus possible ces exigences et les données dont nous disposons généralement à partir des registres comptables. Nous pourrions alors collaborer étroitement avec le nombre relativement restreint d'entreprises qui fournissent des logiciels comptables à un grand nombre de petites et moyennes entreprises au Canada. Nous devrions être en mesure de les aider à créer un ou plusieurs modules de déclaration statistique, qui pourraient permettre de tirer des données des registres comptables des entreprises, de les intégrer à un fichier électronique et de les envoyer à Statistique Canada, grâce à des lignes téléphoniques sans frais, sur une base régulière. Il est possible que certaines de ces suggestions soient mises à l'essai dans les années qui viennent.

L'UTILISATION DE W3 COMME MÉTHODE DE COLLECTE DE DONNÉES : CONCEPTION DES ENQUÊTES DE L'AVENIR

Richard L. Clayton¹

RÉSUMÉ

Internet est devenu le symbole de la Société de l'information. Cet outil technologique d'accès et de partage de l'information s'intègre à tous les secteurs de notre culture. Internet et l'adresse (Universal Resource Locator) de W3 («World Wide Web») sont raccordés à pratiquement tous les autres modes de communication, qu'il s'agisse de quotidiens, de magazines, de télévision ou de radio. On s'attend à ce que cette technologie fournisse aux gens toute information requise sous quelque forme que ce soit, une réponse à toute question, remplaçant une bonne partie de la presse écrite. L'autoroute de l'information a ouvert de riches possibilités de collecte électronique pour les enquêtes d'établissement. Les résultats transformeront éventuellement les enquêtes d'établissement pour en faire le mode d'enquête le plus actuel, le plus exact et le plus rentable.

MOTS CLÉS : Internet; collecte de données; technologie; courrier électronique; W3.

1. INTRODUCTION

Les possibilités d'un accès élargi et instantané par l'entremise d'Internet et de W3 («Word Wide Web») ont déjà poussé les chercheurs à envisager une amélioration des enquêtes. Le présent article passe en revue les caractéristiques actuelles d'Internet/W3 dans le cadre de la collecte de données, examine certaines publications, considère l'élaboration d'un système de collecte de données de type W3 et cerne les avantages et les lacunes du courrier électronique/Internet relativement à d'autres méthodes de collecte de données par enquête assistée par ordinateur (CASIC, Computer Assisted Survey Information Collection), en ce qui concerne la qualité, l'actualité et le coût, et en examine l'application aux enquêtes futures tout en abordant la question de la confidentialité. L'auteur décrit également les premiers résultats d'un projet pilote intégré de courrier électronique/W3 mené par le Bureau of Labor Statistics (BLS) dans le cadre du programme CES (Current Employment Statistics).

1.1 Internet et W3

Internet est un réseau mondial d'ordinateurs reliés par un protocole de communication standard. Ce réseau a été lancé en 1969 par le Department of Defense des États-Unis. Des fonds ont été versés par la Defense Advanced Research Project Agency en vue de l'élaboration d'ARPANET, qui devait servir de réseau polyvalent pour les communications militaires. W3 est une interface graphique d'Internet. W3 est une réalisation de la CERN, laboratoire de physique européen. La première version du logiciel Web a été lancée en 1991. La diffusion de Mosaic en 1993 a rehaussé énormément la popularité de W3, qui représente à l'heure actuelle l'application Internet la plus répandue et la plus prometteuse.

¹ Richard L. Clayton, U.S. Bureau of Labor Statistics, Room 4860, 2 Massachusetts Avenue, N.E., Washington, D.C. 20212.

1.2 Évolution des méthodes CASIC

Au cours des deux dernières décennies, les méthodes de collecte automatisée ont évolué rapidement, depuis CATI (Computer Assisted Telephone Interviewing) dans les années 1970 jusqu'à CAPI, TDE et VR lorsque les micro-ordinateurs ont fait leur apparition. Ces méthodes n'exigeaient guère du répondant que l'accès à un téléphone à clavier.

À la fin des années 1980 sont apparus les questionnaires informatisés auto-administrés (CSAQ, Computerized Self-Administered Questionnaires). Cette méthode, également connue sous le sigle CASI (Computer Assisted Self-Interviewing), a été la première à tirer avantage des nouveaux micro-ordinateurs perfectionnés auxquels avaient accès nos répondants. Les CSAQ permettent à nos répondants de charger le logiciel fourni dans leur OP de façon à saisir et corriger leurs propres données. Les méthodes CSAQ sont donc très semblables à CATI, sauf que le logiciel joue le rôle d'intervieweur. Tout comme CATI, les CSAQ comportent un aiguillage et des vérifications en direct (OMB, 1990).

On a utilisé CASIC en grande partie pour améliorer la qualité des données recueillies et corrigées à la source, tout en contrôlant les erreurs d'intervieweur par l'entremise d'un aiguillage commandé par ordinateur (Nicholls, Baker et Martin, 1995). Ces méthodes ont permis d'améliorer la qualité des données, de rehausser leur actualité et de réduire les coûts. Puisque W3 représente un mode de communication bon marché et facile à utiliser, appuyé par un grand nombre d'ordinateurs portatifs très puissants, il est tout à fait logique, dans le cadre de CASIC, de faire appel désormais à W3 pour la transmission des données.

1.3 Avantages de W3 pour la collecte des données

Souplesse de W3 : Contrairement aux méthodes de collecte téléphonique des années 1980, W3 se prête à tout un éventail d'enquêtes et d'activités connexes. Souvent, les méthodes de collecte téléphonique étaient limitées par la longueur et la complexité du questionnaire, la fréquence du cycle de collecte et la gamme d'activités qu'il était rentable d'entreprendre par voie téléphonique.

Longueur du questionnaire : CATI se limitait généralement à des enquêtes pouvant se dérouler en une séance de quelques minutes. Des problèmes survenaient lorsque le répondant devait consulter ses dossiers. De même, TDE et VR sont limités par le nombre de fois qu'un répondant est prêt à appuyer sur une touche et par le nombre de questions auxquelles il veut bien répondre par machine. Pour sa part, W3 offre la possibilité d'intégrer des questionnaires structurés, peu importe la forme ou la longueur, y compris les modèles de type formulaire ou les concepts traditionnels de questions consécutives. Le répondant est libre de consulter des dossiers aussi souvent qu'il le faut et de répondre à une partie du questionnaire, quitte à y revenir plus tard.

Fréquence des enquêtes : Les enquêtes W3 en cours sont faciles à poursuivre si l'on utilise un fichier de coordonnées qui inclut l'adresse électronique. Les enquêtes de type multimode unique sont plus difficiles à mettre en oeuvre car il n'est pas toujours facile d'obtenir un fichier complet de coordonnées (courrier électronique, courrier, numéro de téléphone).

Modification du questionnaire : W3 offre la même grande souplesse que le courrier traditionnel pour ce qui est de la modification du contenu (l'ajout de nouvelles données, par exemple) ou d'enquêtes complémentaires périodiques. Dans un environnement W3, les répondants ont accès au serveur et utilisent le logiciel qui s'y trouve. Ainsi, il est possible de modifier le système, puis de le charger et maintenir, en un même endroit. Une fois le système chargé, tous les répondants ont immédiatement accès au logiciel modifié.

Améliorations du produit et du service à la clientèle : Les systèmes W3 se prêtent à toute gamme de produits d'enquête ou autres fonctions de facilitation pour les répondants. Les responsables d'enquêtes ont pris soin d'associer l'utilité et l'importance des données publiées à la transmission de microdonnées. Ces efforts ont donné lieu à des livrets, des brochures et des communiqués. L'interface W3 ira encore plus loin en permettant de caractériser les données fournies par chaque répondant, en fonction par exemple de leur industrie, de leur état et de la nation. De plus, des fonctions multimédias pourraient permettre de monter en direct des «coupures» sur l'utilisation des données, qu'il s'agisse de la presse écrite, de la radio ou de la télévision, de façon à rehausser l'utilité et la valeur des efforts de transmission des données.

Exactitude : Pour ce qui est de nos enquêtes, l'exactitude sera également améliorée de différentes façons. Les microdonnées provenant des employeurs se fondent de plus en plus sur des tableaux préparés directement par ordinateur plutôt que sur des sources secondaires. Le répondant peut réagir directement à toute demande de correction.

Actualité : Nos clients bénéficieront de données plus actuelles. Dans le cas de certaines enquêtes, on aura donc rapidement des estimations «définitives», ce qui éliminera le besoin d'estimations «préliminaires» et pour d'autres, réduira l'ampleur des révisions entre les estimations préliminaires et définitives.

1.4 Le programme CES (Current Employment Statistics)

Le CES est une enquête mensuelle menée auprès de quelque 390 000 établissements non agricoles. Puisqu'il assure la collecte d'un faible nombre de données liées à la paye, le CES se prête bien à des méthodes CASIC. Au cours des 14 dernières années, les responsables du CES ont examiné et utilisé les méthodes CATI, TDE, VR (reconnaissance de la parole) et EDI (échange électronique de données) avant de considérer le coût de W3 et ses possibilités de réduction des erreurs. Un aspect qui a poussé le CES à mettre en oeuvre les méthodes CASIC est le très court créneau de collecte des données (2 à 2,5 semaines seulement) pour ce qui est des estimations préliminaires. Dans le cadre de cette gamme de méthodes, on a amélioré de 30 % les taux de réponse moyens des répondants retardataires et on a réduit de 38 % les révisions mensuelles moyennes apportées aux estimations préliminaires. À l'heure actuelle, plus de 230 000 répondants utilisent chaque mois la technologie TDE et l'on s'attend à ce que des milliers d'autres l'adoptent.

2. MÉTHODOLOGIE W3

2.1 Combinaison de deux méthodes CASIC en une seule

L'élaboration d'une méthodologie d'enquête W3 suppose que l'on utilise des outils mis au point dans le cadre de deux autres méthodes. Il s'agit des procédures de contact avec les répondants et des techniques automatisées d'auto-interrogation familières du type CSAQ (Computerized Self-Administered Questionnaires). La combinaison de ces deux fonctions marque l'évolution probable de la méthodologie de collecte W3.

Le cycle W3 de collecte commence par un fichier de contrôle qui contient l'adresse électronique des répondants de même que les coordonnées d'usage (nom, adresse, numéro de téléphone). Le formulaire de collecte est une «page Web» standard comportant une image du questionnaire, des instructions, des définitions et des liens hypertexte pour les définitions. Une adresse électronique est affichée pour permettre de signaler les problèmes et de présenter des demandes. À l'approche du cycle de collecte, le répondant ouvre son courrier électronique et y trouve un rappel, se rend à la page d'accueil CES, accède l'écran de collecte des données et fournit les données requises. Dès que le répondant clique sur l'icône

«submit data», les données sont transmises à l'organisme d'enquête. Les échéances sont enregistrées par voie électronique et à des moments préétablis, un message de rappel de non-réponse est envoyé automatiquement.

Les méthodes TDE existantes éliminent une bonne part des activités fastidieuses d'envoi postal, de réception et de saisie des données. Par contre, aucune méthode n'a encore réglé directement une autre activité dispendieuse, la correction et la réconciliation des données. Les fonctions actuelles de correction et de réconciliation peuvent elles aussi être traitées directement dans le cadre de la collecte W3; comme pour la méthode CSAQ, cela permet au répondant d'examiner directement les erreurs possibles et de les corriger au besoin.

2.2 Méthodes de contact avec les répondants

Dans une enquête postale traditionnelle, il fallait attendre que le répondant reçoive le formulaire pour y répondre. Lorsqu'une réponse n'était pas transmise après un certain temps, un message de non-réponse était envoyé. La méthode TDE élaborée par le BLS intègre ces trois mêmes types de message, échelonnés soigneusement, que l'on peut reprendre dans le cadre d'une méthodologie W3. Les étapes TDE essentielles sont : 1) un préavis sous forme de carte postale ou de télécopie qui remplace le formulaire postal comme stimulus, 2) un appel automatisé à un numéro sans frais afin de fournir des réponses numériques à une série de questions préenregistrées et 3) un message de non-réponse par CATI ou par télécopie (Werking et Clayton, 1994). Dans le cadre de W3, les préavis et les messages de non-réponse ainsi que le choix des délais deviennent des messages électroniques (voir la figure 1 ci-dessous).

Ainsi, la méthodologie tout entière peut être automatisée, depuis les délais et le contenu des messages électroniques, jusqu'aux messages de non-réponse adaptés au destinataire selon la disponibilité des données requises. Puisque la saisie et la correction des données se font en direct, il est possible d'en arriver à une méthodologie vraiment «sans papier et sans intervention humaine» (Werking, 1994).

2.3 La méthode d'interview CSAQ

Les questionnaires CSAQ (Computerized Self-Administered Questionnaires), appelés aussi CASI (Computer Assisted Self-Interviewing), remettent essentiellement l'outil CATI ou CAPI entre les mains du répondant qui lit les questions et fournit lui-même les réponses. La différence, relativement à CATI/CAPI, c'est que l'outil CSAQ doit intégrer toutes les connaissances et la formation de l'intervieweur, y compris les questions supplémentaires. Bien entendu, les outils les plus puissants, l'aiguillage automatisé et les corrections en direct, ont une importance primordiale si l'on veut que cette technique soit plus qu'une simple saisie de données à distance. Les CSAQ comportent deux fonctions clés, le questionnaire lui-même et le véhicule de transmission (p. ex., disquette d'envoi, feuillet d'imprimante ou fichier électronique). Les aspects particuliers des CSAQ qui s'appliquent directement à la collecte W3 sont la conception des écrans et la navigation d'un écran à l'autre, ou d'une page à l'autre.

Figure 1 : Comparaison des types de message pour le répondant : TDE par opposition à W3

Fonction Méthodologique	TDE	W3	Système complet entièrement intégré
Préavis mensuel	Carte postale ou télécopie	Courrier électronique	Production automatique
Transmission des données	Liaison avec le système TDE par numéro 800	https://www.ces.bls.gov	Réception, correction et stockage instantanés des données
Message de non-réponse	Appel téléphonique, télécopie ou carte postale	Courrier électronique	Production automatique en fonction d'un calendrier mensuel

On peut considérer les méthodes W3 comme un prolongement des CSAQ qui leur ressemblent beaucoup et qui offre en plus certains avantages (réception immédiate des données, mise à jour centrale de l'outil de collecte, élimination de l'envoi postal ou d'un téléchargement interminable). Cela suppose une conception W3 qui prévoit un téléchargement rapide et polyvalent d'écrans individuels.

2.4 Aspects de la conception de W3

W3 pose un certain nombre de difficultés aux concepteurs. La sécurité, le bilan des transmissions et la conception de l'interface sont des questions clés. Les enquêtes fédérales doivent respecter rigoureusement la confidentialité des données des participants, ce qui suppose des techniques de codage (voir ci-dessous).

Pour ce qui est du système tel qu'il est présenté aux répondants, il faut prévoir la conception de chaque écran et la navigation d'un écran à l'autre et à l'intérieur d'un même écran. Parmi les spécialistes de l'environnement de base GUI, mentionnons Ben Shneiderman, du Human Computer Interaction Laboratory (Université de Maryland) et, pour ce qui est de la conception des sites W3, Jakob Nielsen, de Sun Microsystems.

La stratégie de Shneiderman consiste à construire des systèmes qui répondent aux besoins de l'utilisateur. En principe, une bonne conception fidélise l'utilisateur pour le plus grand bien du système lui-même. Les interfaces devraient être simples et axées sur l'intuition, et fournir une forme quelconque de réaction ou de récompense (Shneiderman, 1993).

Pour ce qui est de W3, Jakob Nielsen souligne également le caractère intuitif de la conception. Dans la plupart des cas, les utilisateurs ne devraient pas avoir à lire la documentation. Il recommande également «un dialogue» avec les utilisateurs comme élément essentiel du processus d'élaboration (Nielsen, 1996).

Le choix entre un défilement et des «pages» est une première considération. De nombreux sites W3 font appel à une fonction de défilement, mais il s'agit surtout de sites axés sur du texte. Nielsen souligne que l'utilisation de sites axés sur des pages est plus intuitive et moins déroutante. Les répondants peuvent ignorer que d'autres questions sont cachées, et les systèmes axés sur des pages devraient se fonder sur l'aiguillage qui est un aspect courant de CATI, CAPI et CSAQ. Il peut y avoir un compromis entre le nombre de téléchargements et le temps qu'on y consacre, d'une part, et la clarté et le contrôle qu'offre un système axé sur des pages, d'autre part.

Permettre aux répondants de naviguer en connaissance de cause d'un écran à l'autre et à l'intérieur d'un même écran est un élément clé de la satisfaction. L'écran doit expliquer clairement comment se déplacer à l'intérieur du questionnaire.

Le fait d'indiquer la fonction exacte des boutons facilite la conception. De plus, des fonctions familières permettant de reculer ou d'avancer, ou encore des boutons jouant un rôle semblable, aident l'utilisateur à naviguer et devraient être intégrés de façon à faciliter les déplacements. Il est également possible de désactiver certaines fonctions, afin d'empêcher le répondant de se perdre en essayant de découvrir toutes les ramifications, c'est-à-dire de n'ouvrir que certains embranchements.

La composition des pages est également primordiale. Les spécialistes des enquêtes W3 devraient connaître l'environnement des répondants. À l'heure actuelle, les connexions modem standard comportent normalement 14,4 ou 28,8 kb. Par conséquent, afin de minimiser le fardeau des répondants, mesuré selon le temps mis à télécharger la page, on ne devrait pas utiliser beaucoup de graphiques. Les téléchargements interminables sont à éviter.

2.5 Le continuum W3-CSAQ

Une différence essentielle entre W3 et les CSAQ est l'emplacement de l'outil d'enquête automatisé. Les CSAQ placent l'outil dans l'ordinateur personnel du répondant, tandis que W3 sous sa forme idéale offre un accès central à l'instrument. Des variantes de W3 peuvent déplacer certaines fonctions vers l'ordinateur du répondant. Ainsi, puisqu'il est difficile à l'heure actuelle d'assurer un stockage sécuritaire de microdonnées historiques à l'extérieur du logiciel de filtrage en cas de correction longitudinale, le stockage se fait parfois dans l'ordinateur du répondant. Compte tenu de ce déplacement de certaines fonctions entre l'ordinateur du répondant et le serveur W3 central, à un moment donné, le recours à l'ordinateur du répondant fait que l'on a des CSAQ plutôt que W3. Le continuum entre ces deux méthodes doit être abordé par chaque concepteur en fonction des besoins de chaque enquête.

2.6 Méthode de conception globale en direct

Le remplacement éventuel des méthodes traditionnelles par W3 exigera un examen soigné de toute la recherche axée sur le courrier. Le résultat serait un bon point de départ pour la méthodologie W3. Dans le cadre TDE, par exemple, on a atteint des taux de réponse très élevés en combinant des préavis, des interfaces de saisie de données faciles à utiliser et des messages de non-réponse soigneusement programmés. Est-ce que W3 fera de même? La méthode de conception globale (TDM, Total Design Method) offre une stratégie rigoureuse qui maximise les taux de réponse (Dillman, 1978). Suivant cette méthode, chaque fonction de l'enquête (préavis, instrument d'enquête, rappels et choix des délais) peut servir à améliorer les taux de réponse. La figure 2 montre comment certaines fonctions TDM peuvent être mises en œuvre dans W3. La conception de W3 devrait être intuitive et accessible aux répondants. Une bonne conception favorise des taux de réponse élevés.

Figure 2 : Transposition de la méthode de conception globale à W3

Fonction d'enquête	Recommandation TDM	Application W3
Correspondance de l'enquête	La correspondance devrait être personnalisée.	Le préavis et le message électronique de non-réponse s'adressent à un répondant particulier. Le questionnaire est toujours disponible sur le site Web.
Frais postaux	L'enveloppe-retour comporte un timbre.	On pourrait avoir recours au courrier électronique «recommandé» pour le préavis et les messages de non-réponse.
Formulaire d'enquête	Formulaire simplifié : - espace blanc - type livret - définitions annexées	On peut prévoir un écran qui maximise l'espace blanc. Les définitions peuvent être incluses sous forme de liens hypertexte.
Listes de questions	Priorités : un questionnaire convivial court donne des taux de réponse plus élevés qu'un questionnaire convivial plus long.	W3 permet aisément une conception conviviale. La longueur apparente du questionnaire peut être minimisée à l'aide de liens hypertexte.

2.7 Accès des répondants au courrier électronique et à W3

Les employeurs ont réagi aux pressions accrues de compétitivité internationale en réduisant et en nivelant leurs effectifs de façon à accroître la productivité et à mieux maîtriser la structure des salaires et des prix. Toutefois, les employeurs ont également investi davantage dans la technologie des ordinateurs et des communications, au cours des années 1980, ce qui a peut-être été plus important comme moyen d'accroître la productivité, de relier les activités nationales et internationales et d'assurer un accès

immédiat à des informations critiques sur les inventaires, le personnel et les mouvements de trésorerie. En 1991, pour la première fois, les entreprises des États-Unis ont dépensé plus d'argent sur les ordinateurs et les communications que sur l'équipement industriel, minier et agricole et les machines de construction (Werking, 1994). Ces investissements devraient entraîner une plus grande disponibilité de la technologie et un meilleur accès à Internet pour certains groupes de travailleurs.

Dans une enquête de 1995 menée auprès de 404 dirigeants principaux de l'information (DPI) parmi les entreprises de Fortune 2000 (Spanski, 1996), il a été question de plusieurs indicateurs clés des possibilités actuelles et futures du courrier électronique. Or 89 % des DPI possédaient le courrier électronique au sein de leur entreprise. La moitié environ des autres DPI s'attendaient à avoir accès au courrier électronique au cours des deux prochaines années. Cependant, 60 % environ de leurs employés y avaient accès. Moins de la moitié, c'est-à-dire 44 %, avaient accès à Internet. Au début, la plupart de ces grandes entreprises adoptent le courrier électronique comme moyen d'améliorer les communications internes et le processus interne de prise de décision. Les entreprises qui ont accès à Internet estiment que c'est une façon de perfectionner la prise de décisions, ce qui indique que les entreprises sans accès à Internet ne tarderont pas à l'adopter.

On évalue à 14,7 millions le nombre de ménages aux États-Unis qui ont accès à Internet, soit le double du nombre signalé il n'y a qu'un an (Wall Street Journal, 1996). On évalue à 9 millions d'adultes le nombre de personnes y ayant accès chaque jour. Un sondage de Nielsen en 1996 estimait que 30 % des adultes qui travaillent ont accès à Internet (Hoffman, Kalsbeek et Novak, 1996). Le recours à W3 pour les messages transmis à des ménages pourrait poser certaines difficultés. On suppose, par exemple, que les gens retiendront Internet une fois le service mis en place. C'est largement ce qui s'est produit pour le téléphone et les télécopieurs. Toutefois, il y a lieu de croire que certaines personnes abandonnent et adoptent le service Internet à différents moments, et qu'elles pourront changer de fournisseur selon l'évolution des prix et des services, phénomène de «roulement» bien connu dans l'industrie, les adresses électroniques étant dès lors changées. Il est possible que l'évolution, dans l'industrie téléphonique, vers un seul numéro pour la maison, le bureau et l'appareil portatif sera également reflétée dans l'espace cybernétique. Le roulement pourrait être moins élevé au sein des établissements, mais il sera tout aussi difficile de tenir à jour l'adresse électronique des répondants que le numéro de téléphone de ces mêmes répondants. Heureusement, il n'y a eu aucun changement d'adresse électronique parmi les 50 répondants qui communiquent actuellement avec CES à l'aide de W3 (voir ci-dessous).

Cette leur d'espoir dépend des répondants ciblés pour une enquête particulière. Pour un questionnaire destiné à des DPI, on pourrait envisager W3 comme le véhicule de collecte principal. Toutefois, la situation est peut-être moins prometteuse pour la plupart des répondants d'établissement. Dans un sondage non scientifique mené en 1995 auprès de 1 300 répondants à une enquête CES (personnel de la paye typiquement), seulement 6 % des personnes avait accès au courrier électronique. Dans le cadre d'un prototype de W3 lancé en direct en 1996, le CES a trouvé que 10,7 % des répondants qui utilisaient déjà la TDE satisfaisaient aux exigences d'admissibilité de l'accès à Internet (outil de navigation Netscape 2.0 ou mieux) et acceptaient volontiers de participer au projet pilote. Ces unités étaient concentrées dans des secteurs pour lesquels on s'attendait à un taux d'accès élevé, c'est-à-dire des services d'informatique et de traitement des données et des administrations locales et d'État. Il n'y avait pas de répondants qui satisfaisaient à nos exigences dans d'autres industries de service (voir le tableau I).

3. RÉSULTATS DU PROJET PILOTE W3 DU CES

En mars 1996, le CES a lancé un projet pilote de collecte W3. Ce projet doit déterminer la faisabilité et fournir des renseignements sur la conception et l'intégration. L'évolution du système dépendra de notre aptitude à utiliser de nouveaux logiciels pour en établir les fonctions. Nous cherchons à établir un système rudimentaire à l'échelle réduite afin de surmonter les difficultés, au lieu d'attendre qu'un système complet

puisse être mis au point. Le formulaire CES axé sur le courrier permet de recueillir cinq ou six unités de données chaque mois d'un nombre fixe de répondants. Les données mensuelles sont inscrites sur des lignes consécutives couvrant une année entière dans une série de rangées et de colonnes. Les entêtes de colonne indiquent les données requises.

Le projet pilote W3 ne reprend cette rangée de données que pour le mois courant. On ne montre pas plusieurs mois à cause de la difficulté, à l'heure actuelle, de garantir la sécurité de microdonnées rapportées antérieurement. Les entêtes de colonne figurent sous forme de liens hypertexte renvoyant à des définitions des données. Pour avoir accès au système, le répondant doit inscrire son numéro unique et son mot de passe avant de passer à une page d'accueil qui affiche les énoncés exigés par l'Office of Management and Budget au sujet de la confidentialité et des obligations du répondant. Le numéro unique et le mot de passe servent également à identifier le formulaire exact à présenter au répondant. Le répondant inscrit le numéro du mois de référence et ses données. Cette étape terminée, il s'agit d'appuyer sur un bouton étiqueté «Send data to BLS» qui assure le codage des données et leur transmission au serveur BLS. À l'heure actuelle, il existe un module prototype de correction interactive et il sera mis à la disposition des répondants W3 d'ici quelques mois.

Présentement, 52 répondants CES signalent leurs effectifs, leur feuille de paye, ainsi que les heures de travail à l'aide de W3. Tous transmettaient auparavant leurs données par TDE et connaissent le cycle de collecte mensuel. Chaque mois, juste avant le temps prévu pour la collecte, un message est transmis par courrier électronique à titre de préavis, tout comme des cartes postales et télécopies d'avis sont transmises aux répondants TDE. Les personnes qui ne répondent pas dans les délais prévus reçoivent un message de non-réponse par courrier électronique.

Cette recherche a donné deux résultats très importants. Tout d'abord, les taux de réponse de base pour nos estimations préliminaires relativement à ces unités sont les mêmes, c'est-à-dire 76 %, pour W3 qu'ils l'étaient pour TDE. Nous ne voyons donc pas pourquoi la collecte W3 ne pourrait pas assurer les mêmes taux de réponse élevés observés pour des méthodes CASIC semblables d'auto-évaluation par téléphone : TDE et VR. Deuxièmement, les messages par courrier électronique semblent tout aussi efficaces que d'autres méthodes. Le nombre d'unités W3 nécessitant un message de non-réponse par courrier électronique, qui est de 25 % environ, est presque le même pour l'échantillon TDE global. Ainsi, l'ensemble de messages décrit dans le tableau 1 donne des résultats efficaces comparables à ceux des méthodes TDE fructueuses.

Ces deux aspects importants de la méthodologie ouvrent la voie à des essais plus poussés et à des améliorations, y compris la présentation du courrier électronique sous forme de fichiers sonores ou de graphiques.

Tableau 1 : Résultats de la collecte type W3 du CES : de mars à septembre 1996

	Total actuel	Services d'informatique et de traitement des données (CTI=737) n=313	Autres CTI de service n=121	Administration locale ou d'État n=264
Courrier électronique seulement	4,1 %	6,4 %	0 %	2,4 %
L'entreprise dispose de fonctions, le répondant ne s'en sert pas	10,5 %	12,8 %	7,0 %	8,7 %
Outil de navigation ancien ou incompatible	4,5 %	6,4 %	5,8 %	0,6 %
Courrier électronique et Web non montés dans l'ordinateur	2,1 %	3,4 %	0 %	1,2 %
<i>Outil de navigation compatible, courrier électronique/Web monté dans l'ordinateur</i>	<i>10,7 %</i>	<i>14,3 %</i>	<i>0 %</i>	<i>10,4 %</i>
TDE préférée	2,1 %	3,4 %	0 %	1,2 %
Pas de fonction courrier électronique ni W3	62,5 %	49,4 %	79,1 %	74,9 %
Plus en affaires/non pertinent	3,5 %	3,9 %	8,1 %	0,6 %
Total	100 %	100 %	100 %	100 %

La pénétration assez faible de W3 dans les ménages aussi bien que dans les établissements, malgré sa croissance, indique nettement qu'il va y avoir inévitablement des modes de collecte mixtes au cours des années à venir. Le processus d'élaboration et d'intégration de deux modes ou plus nécessitera plusieurs systèmes de contrôle et rendra la poursuite et l'intégration plus difficiles, accaparant le soutien existant et entraînant possiblement l'oubli de détails importants.

4. COÛTS

Depuis des dizaines d'années, nous avons investi d'importantes sommes d'argent dans l'élaboration et le perfectionnement d'activités centralisées et décentralisées à forte main-d'oeuvre afin d'assurer la qualité de nos estimations. Ces activités englobent la collecte et le contrôle de la collecte, des étapes multiples et un suivi de non-réponse, la saisie et sa vérification, ainsi que la vérification et la correction. Toutefois, dans un cadre de transmission W3 toutes les activités de collecte peuvent être entièrement automatisées et centralisées sur un réseau local spécialisé. Les messages sont transmis par voie électronique à des dates préétablies et l'information est dûment saisie. Les vérifications sont mises en oeuvre dans le cadre de la séance de collecte de données W3.

Un des faits saillants du système CASIC est un changement de la structure des coûts en vertu duquel les démarches postales traditionnelles à forte main-d'oeuvre sont remplacées par d'autres facteurs à forte intensité de capital. Les coûts de la main-d'oeuvre ont tendance à augmenter tandis que les coûts de la technologie ont tendance à diminuer avec le temps. Dans le cadre traditionnel d'un envoi postal, le personnel de bureau ou semi-professionnel se charge de préparer et de remplir les enveloppes, au moment de l'envoi, puis de dépouiller les questionnaires retournés, et d'assurer la précorrection et la saisie. CATI comporte une fonction courrier réduite et de nouveaux coûts pour la technologie, mais retient des frais

élevés de main-d'œuvre pour les intervieweurs. Dans le cadre TDE, la plupart des activités à forte main-d'œuvre sont remplacées par l'intervention du répondant qui inscrit les données, et ce sont les coûts téléphoniques qui définissent les coûts unitaires globaux. La collecte W3 offre une saisie et une transmission des données pratiquement gratuites lorsque les frais d'accès à Internet pour l'organisme d'enquête sont absorbés par l'organisme et un grand nombre de répondants. Le prix de ce genre d'accès est actuellement à la baisse compte tenu de la concurrence.

Tableau 2 : Coûts unitaires mensuels pour la transmission des données (CES)

Fonction	Courrier - timbres	TDE/télécopie	Internet/W3
Sortie	0,32 \$	0,08 \$	0,00 \$
Entrée	0,32 \$	0,16 \$	0,00 \$
Message de non-réponse	0,10 \$	0,04 \$	0,00 \$
Total	0,74 \$	0,28 \$	0,00 \$

Pour toutes les enquêtes sauf les plus petites, la disparité croissante entre les frais postaux et téléphoniques favorisera l'adoption de CASIC à mesure que la réduction des frais de transmission compense les coûts de mise en oeuvre et d'entretien des systèmes. Les coûts unitaires de transmission des données que contient le tableau 2 montrent bien la forte réduction des coûts que permet une transition de TDE à Internet pour la collecte. Dans le cas des organismes qui achètent un accès illimité à W3, le coût moyen d'une séance devrait se rapprocher de 0.

5. EFFETS ORGANISATIONNELS

Comme ils l'ont fait depuis deux décennies, nos organismes continueront d'évoluer à mesure que d'importantes améliorations technologiques sont mises en oeuvre. Comme nous l'avons vu, l'adoption élargie des méthodes CASIC et, en particulier, la collecte W3, entraîneront la transmission de données corrigées directement à l'organisme d'enquête. Ainsi, l'attribution de ressources en main-d'œuvre et autres se déplacera vers d'autres facteurs et de nouveaux éléments de production. L'idée que la collecte W3 offre une méthode «sans papier et sans intervention humaine» (Werking, 1994) comporte de profondes répercussions sur nos organismes. Même si l'utilisation généralisée de W3 pour la collecte en est encore à l'état embryonnaire, nous sentons déjà un déplacement des ressources vers d'autres méthodes.

Il est certain que des méthodes comme TDE et VR reposent sur la saisie des données par le répondant et sur un recours moindre à des intervieweurs.

Les paramètres organisationnels particuliers d'une enquête donnée dépendront de la place qu'elle occupe sur l'échelle CASIC. Les enquêtes menées par courrier seront remaniées de fond en comble, tandis que les enquêtes qui font appel actuellement à CATI ou TDE subiront des effets moindres dans des catégories de ressources différentes. Ainsi, dans le cas d'une enquête axée sur le courrier ou TDE, le travail de correction sera réduit appréciablement. Les enquêtes qui font déjà appel à CATI observeront une baisse des coûts de main-d'œuvre. Dans le cas des enquêtes par TDE, la plupart des économies de main-d'œuvre ont été réalisées en termes d'envoi et de saisie, et les frais de vérification et de téléphone sont appelés à disparaître ou presque.

De nouvelles ressources, tirées probablement d'une réduction des coûts d'entrevue, seront accordées aux nouvelles démarches d'élaboration et d'entretien des systèmes. Il faudra prévoir de nouvelles dépenses pour la gestion des réseaux, des serveurs nouveaux et modifiés, ainsi que le logiciel, y compris le codage. Il est tout aussi important de savoir investir dans l'infrastructure qui sous-tend les nouvelles méthodes d'enquête que de réaliser soigneusement toutes les économies possibles se rapportant aux procédés. Le rôle des concepteurs de questionnaires prendra de l'ampleur, surtout pour ceux qui font la transition à partir des enquêtes sur papier.

6. LA SÉCURITÉ

L'aspect technologique sans doute le plus critique de l'infrastructure Internet est la sécurité des données transmises et mises en mémoire. Cette restriction est signalée par tous les adeptes du Web et elle a suscité l'intérêt d'une bonne partie de la collectivité informatique. Certaines caractéristiques du profil de sécurité Web sont les suivantes :

- Authentification du répondant
- Protection contre l'espionnage durant la transmission (sécurité des données par paquets)
- Protection de la séance (détournement)
- Protection de données confidentielles arrivant au serveur
- Protection contre l'accès de non-usagers au réseau local de l'organisme

Ce domaine évolue rapidement et de nouveaux outils et logiciels viendront modifier nos stratégies actuelles. Ces questions doivent être abordées par tous les systèmes et procédés W3. La question la plus difficile, qui reste sans solution dans le projet pilote du CES, est le stockage de données historiques en vue d'une correction longitudinale à l'extérieur du logiciel de filtrage. Même en présence de codage, des personnes suffisamment motivées pourraient tôt ou tard réussir un décodage.

Les questions de sécurité liées à la collecte W3 semblent souvent beaucoup plus épineuses que pour nos autres méthodes CASIC. Il y a deux différences : Internet est offert actuellement aux personnes qui désirent explorer et aux curieux qui s'intéressent à la technologie; Internet est un phénomène trop récent pour qu'il existe de la jurisprudence et des dispositions législatives pouvant restreindre le comportement des gens. Le système postal comporte des dispositions législatives qui interdisent l'interférence avec le courrier, de la jurisprudence qui appuie ces dispositions et même un service de police spécialisé dans ce genre de crime. Le système téléphonique comporte également de la jurisprudence et des dispositions législatives qui interdisent l'écoute clandestine. Pour le courrier aussi bien que le système téléphonique, le grand public est au courant de ces lois et les activités criminelles sont signalées périodiquement. Il existe aux États-Unis des lois qui pourraient s'appliquer à Internet. L'Electronic Communications Privacy Act de 1986 et la Computer Fraud and Abuse Act de 1986 s'appliquent peut-être toutes deux à Internet. Lorsqu'une décision officielle aura été prise au sujet de l'application de ces lois, elle pourra figurer sur les pages d'accueil à titre dissuasif.

7. L'AVENIR

À quand l'intervieweur virtuel? : Les adeptes des méthodes CASIC ont tous leur propre vision de l'avenir des méthodes et des technologies, mais il est tout probable que les fascinantes recherches des deux dernières décennies vont restructurer nos procédures d'enquête. Les fonctions W3 qui seront préparées et mises à l'essai dépendront du caractère particulier de chaque enquête. On voit apparaître de nouvelles technologies qui offrent de fascinantes possibilités, y compris les techniques multimédias, sonores et vidéo. Il serait ainsi possible d'envoyer des messages électroniques à complément sonore. Lorsqu'on clique sur l'icône, on pourrait entendre comme préavis ou comme message de non-réponse «veuillez transmettre vos données d'ici le 28». Il serait possible d'inclure des coupures de radio ou de télévision illustrant l'utilisation des données. Il existe actuellement des caméras bon marché que l'on peut coupler à un ordinateur personnel pour des vidéoconférences en direct. Il est concevable que ce genre de caméra fasse partie des emballages de vente comme on le fait pour les disques compacts. Il serait possible aussi d'utiliser ces mêmes caméras dans les enquêtes pour le «service d'assistance en direct». Lorsqu'un répondant passe trop de temps sur une même question, un écran instantané pourrait lui offrir de plus amples renseignements ou des exemples. L'écran instantané pourrait même afficher un icône sur lequel il suffirait de cliquer pour établir une liaison avec un intervieweur dont la photo et la voix apparaîtraient à l'écran du répondant. L'intervieweur aurait devant lui le même écran et d'autres informations historiques

sur le répondant, et serait en mesure d'éclaircir certains points ou encore d'expliquer l'utilisation du système. Ce genre d'entrevue aiderait à surmonter les problèmes.

Le rôle de l'intervieweur continuera d'évoluer en fonction des méthodes CASIC. W3 est assez puissant et convivial pour remplacer l'intervieweur et offrir au répondant toute l'interface humaine rendue possible par l'espace cybernétique. Compte tenu de la technologie déjà disponible, on peut prévoir un «intervieweur virtuel» monté sur des bandes vidéo/audio préenregistrées comme on le fait actuellement pour les jeux interactifs. On réserverait une partie de l'écran pour une image de l'intervieweur abordant les questions dans une langue quelconque, la question ou le formulaire étant également affiché. Un aiguillage axé sur les réponses (au clavier ou par reconnaissance de la parole) assurerait la présentation des séquences visuelles et sonores. Des réponses introduites manuellement ou par la voix orienteraient l'aiguillage et des menus offriraient au répondant des renseignements supplémentaires à différents niveaux. L'intervieweur virtuel pourrait être mis en oeuvre dès maintenant à l'aide de disques compacts interactifs, comme ceux dont se servent nos enfants pour apprendre ou anéantir tous ces gros méchants.

Élaboration de questionnaires en direct et dépannage : Cette vision offre aux responsables d'enquêtes la meilleure fonction imaginable pour améliorer la surveillance et les questionnaires. La surveillance peut se faire par saisie de la frappe ou à distance en observant simplement le déroulement des réponses. L'intervieweur virtuel aura déjà en mémoire des messages sonores et visuels. Dès lors, le système W3 fournirait des messages sonores et des images de l'intervieweur à l'intérieur d'une petite fenêtre dans un coin de l'écran. L'intervieweur pourrait parler ou répondre à des questions posées de vive voix ou inscrites au clavier.

Internet II : La version actuelle d'Internet ne permet pas la vision décrite ci-dessus. Toutefois, la prochaine génération, c'est-à-dire Internet II, est en voie de préparation et les discussions sont axées justement sur le type de fonction qui rendrait possible l'intervieweur virtuel.

8. CONCLUSION

La collecte massive et pleinement automatisée de données par W3 est un phénomène inévitable. Les recherches de base menées jusqu'à présent, bien que ralenties par le caractère nouveau de la technologie, s'appuient sur les résultats de deux décennies de recherches CASIC. Les aspects méthodologiques, qui s'appuient eux aussi sur les recherches et les applications CASIC antérieures, laissent entrevoir une qualité égale ou plus grande des données relativement à d'autres méthodes, une actualité plus grande relativement aux méthodes traditionnelles et égale comparativement à CATI et à TDE, à un coût beaucoup moindre.

Ainsi, l'avenir est assez clair tout comme il est évident que nos enquêtes postales appartiennent au passé. L'autoroute de l'information a ouvert la voie à de nombreuses possibilités de collecte électronique pour les enquêtes d'établissement. Grâce aux résultats de la recherche, les enquêtes d'établissement et de ménage deviendront à la fin les activités d'enquête les plus actuelles, les plus exactes et les plus rentables.

Remerciements : L'auteur tient à remercier Louis J. Harrell et Christopher Manning de leur superbe travail de mise au point du prototype W3 du CES et d'analyse des résultats.

9. BIBLIOGRAPHIE

- Dillman, D. A. (1978), "Mail and Telephone Surveys: The Total Design Method", New York: Wiley-Interscience.
- Hoffman, D. L., W. D. Kalsbeek, et T. Novak. "Internet Use in the United States: 1995 Baseline Estimates and Preliminary Market Segments". Project 2000 Working Paper, Owen Graduate School of Management, Université Vanderbilt, 12 avril, 1996.
- Nielsen, Jakob (1996), "Usability Testing of WWW Designs", <http://www.sun.com/sun-on-net/udesign/usabilitytest.html>.
- Spanski, R., et L. Wickham, (1995), "Connecting the Workplace: Electronic Commerce in Business and Government", Study 944013, Louis Harris and Associates, Inc.
- Shneiderman, Ben (1993), "Designing the User Interface, Strategies for Effective Human-Computer Interaction", Addison Wesley.
- Statistical Policy Working Paper 19 (1990); "Computer Assisted Survey Information Collection", Office of Management and Budget.
- Wall Street Journal, "U.S. Households With Internet Access Doubled to 14.7 Million in Past Year", 11 octobre, 1996, pp. B11.
- Werking, George S. (1994), "Designing the Survey Operations of the Future: A Paperless and People-less Collection Environment." Proceeding of the American Statistical Association Joint Statistical Meetings, Invited Panel on the Future of Establishment Surveys, à l'impression.
- Werking, G.S. et R.L. Clayton (1995); "Automated Telephone Methods for Business Surveys", *Business Survey Methods*, Wiley pp. 317-338.

SESSION 6

ERREURS DE RÉPONSE

L'APPARIEMENT DE MICRO-DONNÉES: UN OUTIL D'ÉVALUATION ET D'AMÉLIORATION DE LA QUALITÉ DES DONNÉES D'ENQUÊTES

L. Gervais-Simard¹

RÉSUMÉ

Cet article relate les expériences récentes de la section de la coordination des enquêtes et de l'amélioration de la qualité en matière d'appariement de micro-données. Trois types d'appariements distincts faits entre des données d'enquêtes de Statistique Canada et celles provenant de dossiers administratifs de Revenu Canada y sont traités. Le but de ces études était d'évaluer le niveau de qualité des données d'enquêtes, d'identifier les erreurs non dues à l'échantillonnage et par la suite d'établir des mesures pour améliorer la qualité des données. Tout ceci afin de rehausser le niveau de qualité des estimations d'emploi et de rémunération des enquêtes menées par la division du travail.

MOTS CLÉS: Données administratives; Revenu Canada; qualité des données d'enquêtes; appariement de micro-données; mesures d'amélioration de la qualité.

1. INTRODUCTION

1.1 Mise en contexte

En 1993, Revenu Canada ajoutait au relevé de compte à la source (formulaire PD) des entreprises deux questions, l'une sur l'emploi et l'autre sur la rémunération brute, et ce à la demande de Statistique Canada. Dès 1994, la division du travail introduisait dans la composition de ses estimations mensuelles les données provenant de ces deux questions, données que l'on qualifie d'administratives. Ce nouveau processus a continué de s'intensifier faisant ainsi croître la portion de données administratives à plus de soixante-dix pour-cent des estimations de l'enquête sur l'emploi, la rémunération et les heures (l'EERH). Comme la comparaison avait déjà été faite, entre les données administratives provenant des formulaires PD agrégées et celles provenant d'enquêtes, nous connaissions donc leur apport positif sur la qualité des estimations d'emploi et de rémunération. Partant de cette prémisse, l'appariement de micro-données entre ces deux sources de données semblait être un outil approprié pour: évaluer la qualité des données d'enquêtes, identifier les erreurs non dues à l'échantillonnage et établir des mesures d'amélioration de la qualité des données d'enquêtes. Comme les données administratives fournissent des informations restreintes, elles ne remplaceront jamais complètement les données provenant d'enquêtes. Des enquêtes plus limitées serviront toujours à attribuer les caractéristiques manquantes aux données administratives.

La section de la coordination des enquêtes et de l'amélioration de la qualité (CEAQ) a procédé au cours de la dernière année à trois expériences distinctes d'appariement de micro-données. Ce sont:

- Les données annuelles de l'enquête sur l'emploi, la rémunération et les heures avec les données du sommaire de la rémunération payée (T-4) de Revenu Canada.

¹ L. Gervais-Simard, Division du travail, Statistique Canada, Ottawa (Ontario), K1A 0T6.

- Les données mensuelles de l'enquête sur l'emploi, la rémunération et les heures avec les données du relevé de compte à la source PD7A(TM)² de Revenu Canada.
- Les données mensuelles de l'enquête sur la rémunération des petites entreprises avec les données du relevé de compte à la source PD7A³ de Revenu Canada.

2. EXPÉRIENCES D'APPARIEMENTS DE MICRO-DONNÉES

2.1 Appariement des données annuelles (1993) de l'enquête sur l'emploi, la rémunération et les heures⁴ (l'EERH) avec les données du sommaire de la rémunération payée (formulaire T-4 Supplémentaire)⁵ de Revenu Canada

2.1.1 Information de base

La qualité des données à évaluer était celle de l'enquête sur l'emploi, la rémunération et les heures, laquelle se tenait tous les mois auprès d'environ 43 000 entreprises en 1993. L'appariement des données de cette enquête mensuelle avec celles de l'état de la rémunération payée (formulaire T-4 supplémentaire) de Revenu Canada s'est fait pour l'année 1993. Les employeurs sont tenus de déclarer les renseignements sur la rémunération payée à Revenu Canada une fois par année, et dans ce cas-ci il s'agit du 28 février 1994. Comme la présente étude s'est tenue à l'été 1995, les données administratives les plus récentes disponibles étaient celles de l'année 1993. Le délai entre la période référence et celle de notre étude est attribuable au traitement que Revenu Canada doit faire à ces données.

Comme il n'y a pas de liens directs entre les entreprises enquêtées par l'EERH et les numéros de comptes à la source de Revenu Canada, on a dû utiliser le registre des entreprises de Statistique Canada afin d'en établir un entre ces deux sources de données. L'EERH étant une enquête-entreprises⁶, cette entité fut donc choisie comme unité de base pour l'appariement des micro-données; le numéro de compte à la source (PD) étant toutefois le dénominateur commun.

2.1.2 Buts et contraintes

Les buts de cette étude étaient de servir à:

- connaître l'écart réel entre les deux sources de données,
- trouver les raisons causant les écarts en contactant les répondants de l'EERH,
- remédier aux faiblesses de l'enquête, par la mise en place de mesures de la qualité.

Au départ, notre équipe avait reçu la directive de procéder à l'appariement de données pour lesquelles des écarts au niveau des données d'emploi avaient déjà été notés entre l'EERH et l'enquête sur la population active (EPA). Pour répondre à cette directive, les entreprises du secteur biens non durables furent ciblées aux fins de cette étude. La seconde contrainte en était une de temps. On a donc dû se restreindre à sélectionner les entreprises identifiées comme ne possédant qu'un numéro de PD, réduisant ainsi le nombre d'entreprises à contacter.

² Le terme PD7A(TM) signifie que l'entreprise effectue plus qu'une remise monétaire par mois à Revenu Canada

³ Le terme PD7A signifie que l'entreprise n'effectue qu'une remise monétaire par mois à Revenu Canada

⁴ Voir Sources et définitions à la fin de l'article.

⁵ Voir Sources et définitions à la fin de l'article.

⁶ Voir définition d'entreprise à la fin de l'article.

2.1.3 Univers

On a donc identifié 1 246 entreprises du secteur des biens non durables répondant à la condition ci-haut mentionnée soit: 1 entreprise = 1 numéro de PD = 1 ou plusieurs établissements⁷. Ces entreprises représentent plus de 30 % de toutes les entreprises (4 066) de l'EERH pour lesquelles au moins un établissement opérait dans ce secteur d'activité économique.

Pour fins d'appariement de micro-données, certaines entreprises ont dû être retranchées soient:

- toutes les entreprises pour lesquelles il n'y a pas de données pour tous les mois de l'année 1993. Il s'agit en fait des entreprises pour lesquelles un ou plusieurs établissements sont entrés ou sortis de l'échantillon de l'EERH au cours de l'année;
- toutes les entreprises pour lesquelles au moins un mois de données a été imputé pour un ou plusieurs établissements, dans le cadre de l'EERH au cours des 12 mois de 1993;
- et finalement, toutes les entreprises pour lesquelles au moins un établissement ne fait pas partie du groupe industriel 37 (biens non durables).

Suite à ces exclusions 99 entreprises ont été retenues pour fin d'appariement des micro-données. Deux variables communes aux deux sources d'information ont été analysées, soient: le nombre de salariés⁸ et la rémunération brute⁹.

2.1.4 L'appariement des variables

L'EERH étant une enquête mensuelle, le nombre maximal de salariés déclarés au cours de l'année pour chacune des entreprises a été utilisé pour représenter le nombre de salariés de 1993. Du côté de Revenu Canada c'est le nombre de numéros uniques d'assurance sociale déclarés par chaque entreprise qui constitue le nombre de salariés. Par conséquent, on ne peut s'attendre à obtenir un résultat identique entre les deux sources de données, tout au plus à des résultats relativement près.

Les données de la variable rémunération brute annuelle, qu'elles proviennent de l'EERH ou de Revenu Canada, devraient correspondre de façon identique s'il n'y avait aucun problème relié à la qualité des données d'enquêtes. Dans le guide de réponse, il est clairement spécifié que le montant de la rémunération brute devrait correspondre au total de tous les montants qui apparaissent dans la case 14 du feuillet T-4. Pour procéder à l'appariement de cette variable, les données d'enquêtes ont été cumulées pour les 12 mois de l'année 1993.

2.1.5 Écart brut suite à l'appariement

À ce stade-ci, l'appariement des 99 entreprises retenues montre l'écart brut entre les deux sources de données pour chacune des variables. L'écart réel sera calculé à la fin de l'exercice, soit après avoir enquêté chacune des entreprises. Pour l'instant, ces résultats sont non significatifs.

⁷ Voir définition d'établissement à la fin de l'article.

⁸ Pour la définition de salarié voir Sources et définitions à la fin de l'article.

⁹ Pour la définition de rémunération brute voir Sources et définitions à la fin de l'article.

Sources	Salariés	Rémunération Brute (000) \$
Feuillet T-4 (Revenu Canada)	38 834	636 326 \$
L'EERH (Statistique Canada)	30 389	653 549 \$
Écart global	8 445 ou 27,8%	- 17 222 \$ ou -2,6%

2.1.6 Méthode d'enquête et questionnaire

Les répondants de l'EERH ont été rejoints par téléphone afin de répondre à un questionnaire¹⁰ sur l'écart des données entre les deux sources. Avant d'aborder ce sujet, l'intervieweur devait vérifier si le lien entre l'entreprise et le numéro de PD était unique, s'il ne l'était pas, ces entreprises étaient exclues de l'étude. Après vérification du lien, l'intervieweur expliquait qu'en procédant à une comparaison entre les données déclarées à Statistique Canada et celles fournies à Revenu Canada, des écarts, qui quelquefois étaient très considérables, avaient été découverts. Le répondant était alors informé de l'ampleur de l'écart trouvé pour son entreprise et ce pour les deux variables communes soit: le nombre de salariés et de rémunération brute. Suite à quoi, des questions permettant de déceler les causes de la sous-réponse ou de la surréponse au questionnaire de l'EERH, étaient alors posées.

2.1.7 Observations suite au contact avec les répondants

L'appariement boíteux des micro-données de la variable salarié a pu être élucidé lors du contact avec les répondants. En effet, ceux-ci ont expliqué l'écart entre les données de cette variable de la façon suivante: il s'agit d'un secteur économique où la rotation de personnel est très élevée. En effet, les bas salaires enregistrés pour les industries du textile et de l'alimentation accentuent ce mouvement de rotation du personnel. Comme chaque personne possède un numéro d'assurance social unique, et que c'est le total de ceux-ci qui constitue le nombre de salariés pour la source administrative, il est donc normal que ce nombre soit de beaucoup supérieur (27,8 %) au nombre de salariés déclarés à l'EERH. Par conséquent il était donc impossible d'établir quel était l'écart réel pour le nombre de salariés entre les deux sources de données. L'analyse pour cette variable ne peut donc être poursuivie.

2.1.8 Raisons expliquant l'écart brut de la variable rémunération brute

Après que les répondants aient fourni des explications sur les causes des écarts entre la rémunération brute déclarée à Statistique Canada et celle fournie à Revenu Canada, on était donc en mesure d'éliminer les entreprises qui faussaient les résultats de l'appariement. En effet, huit entreprises ont dû être retranchées parce que le lien entre l'entreprise et le numéro de PD n'était pas unique. Les renseignements de la base de données du registre central (BDRC) utilisés lors de la sélection des entreprises n'étaient donc pas exacts pour celles-ci. De plus, deux autres entreprises ont été retirées à cause des concepts de l'EERH. Ces numéros de PD incluaient des salariés dont l'activité ne fait pas partie de la population cible de l'EERH, soient des pêcheurs. Donc, si on enlève les écarts représentés par les dix entreprises ci-haut mentionnées, on arrive à un écart net, qui indique un résultat positif de 7,8 % **contrairement à l'écart brut qui était de -2,7 %**. Un résultat positif indique que le montant de rémunération brut provenant de l'appariement de micro données du secteur des biens non durables, est supérieur du côté de Revenu Canada.

L'écart net résulte d'un amalgame des 89 entreprises restantes pour lesquelles des écarts positifs et négatifs ont été enregistrés. Le tableau suivant montre de quelle manière elles sont réparties:

¹⁰ Le questionnaire est disponible en rejoignant l'auteur de l'article.

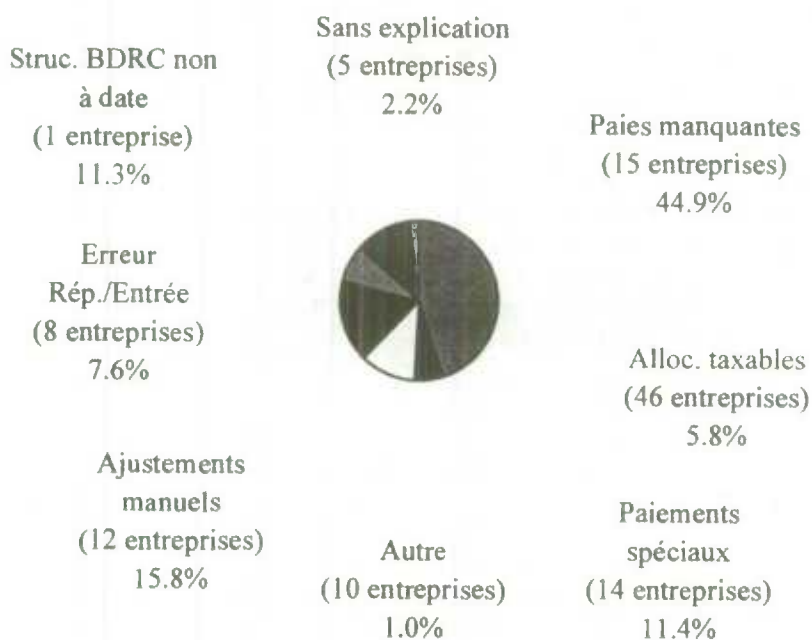
Répartition de l'écart net	(000) \$	Nombre d'entreprises
Partie positive de l'écart net	48 692 \$ ou différence de 9,0%	70 entreprises
Partie négative de l'écart net	- 3 882 \$ ou différence de - 5,4%	19 entreprises
Écart net	44 809 \$ ou différence de 7,8 %	89 entreprises

Le graphique suivant montre séparément les raisons de la partie positive de l'écart net (70 entreprises) suite à l'appariement des entreprises du secteur des biens non durables.

Graphique 1

Distribution de la partie positive de l'ÉCART NET, selon la raison
Rémunération brute 1993
(Secteur des biens non-durables)

Écart \$48,8 millions ou 9.0%
(70 entreprises)



Les raisons montrées au **graphique 1** sont expliquées ici:

Paies manquantes : ce sont les paies des cadres, du personnel administratif ainsi que des paies dites confidentielles que les répondants ont avoué ne pas déclarer à l'enquête sur l'emploi, les gains et la rémunération. Il est facile pour les répondants de ne pas inclure ces types de paies, car elles sont souvent traitées séparément et par conséquent elles sont oubliées.

Ajustements manuels : cette part de l'écart est strictement attribuable à des ajustements faits soit à la phase de vérification ou d'analyse de l'enquête. Le répondant n'est nullement responsable de cette situation.

Paiements spéciaux : il s'agit de sommes qui sont versées aux salariés pour le travail accompli ou pour d'autres versements qui ne portent pas exclusivement sur la dernière période de paye du mois. Ces paiements sont effectués à un moment quelconque au cours du mois et surtout ils ne font pas partie des salaires réguliers. Les répondants ont tendance à ne pas les inclure vu la nature de ce type de paiements. Il y a aussi quelques fois des problèmes reliés soit à l'accessibilité ou la disponibilité de l'information.

Structures de la base de données du registre central (BDRC), non à jour : l'information sur les structures d'entreprises n'étant pas à jour, les données de l'EERH en ont donc subi le contrecoup. En effet, si certains établissements ne sont pas rattachés à une entreprise sur la BDRC, ils ne peuvent être enquêtés.

Erreur du répondant ou d'entrée des données : on a attribué l'écart à ces raisons lorsque les données non corrigées de l'EERH ne correspondent pas aux données dont dispose le répondant dans ses registres.

Allocations taxables : il s'agit d'allocations et d'avantages fédéraux imposables reconnus par Revenu Canada.

Sans explications : ni le répondant ni l'intervieweur ne furent en mesure d'établir une raison plausible à l'écart. Il s'agissait d'entreprises où les écarts étaient minimes.

Incapable de rejoindre le répondant : le numéro de téléphone n'existe plus laissant supposer que l'entreprise est fermée.

Autre : sont comprises ici les entreprises qui ont refusé de répondre aux questions et celles qui avaient procédé à des ajustements annuels, lesquels n'avaient pas été déclarés à l'EERH.

Pour ce qui est de la partie négative de l'écart net, laquelle regroupe 19 entreprises, on note que 77,4 % de celle-ci est causée par des ajustements manuels effectués au cours de l'enquête.

2.1.9 Actions immédiates d'amélioration de la qualité

Lors des interviews, les répondants qui ne déclaraient pas toute l'information requise étaient incités à le faire dès que possible. De plus, l'importance d'une bonne qualité de réponse pour Statistique Canada leur était expliqué. Lorsque de nouvelles personnes ressources étaient identifiées, leurs noms étaient alors référés à la division des opérations et de l'intégration afin que celle-ci les intègre au fichier maître de l'enquête. De plus, le Registre des entreprises était informé de toutes anomalies d'ordre structurelles trouvées lors de l'interview.

Des mesures de qualité plus globales, ayant un impact à plus long terme, ont aussi été élaborées suite aux deux premières études d'appariement de micro-données décrites dans cet article; on les retrouve à la section 2.3 de l'article.

2.2 Appariement des données mensuelles de l'enquête sur l'emploi, la rémunération et les heures (l'EERH) avec le relevé de compte à la source de Revenu Canada, soit les formulaires PD7A(TM)

2.2.1 Information de base

Dans le cadre du remaniement de l'enquête sur l'emploi, la rémunération et les heures, le groupe de l'amélioration de la qualité a été mis à contribution, en contactant des entreprises pour lesquels des écarts majeurs et incompréhensibles avaient été notés lors de l'appariement de micro-données. La division des méthodes d'enquêtes-entreprises (DMEE), en collaboration avec le groupe de l'amélioration

de la qualité, avaient procédé à l'appariement de micro-données de l'EERH avec celles du relevé de compte à la source (formulaire PD7A(TM)) de Revenu Canada pour le mois de mai 1995. Le but premier de cet appariement était d'expérimenter la pertinence des règles d'estimations du nombre de salariés à partir des données administratives, règles qui avaient été élaborées par la division des méthodes d'enquêtes-entreprises. Ces règles permettaient d'obtenir un nombre unique de salariés et de rémunération brute vu que plusieurs remises sont effectuées pour chaque PD à Revenu Canada tous les mois. Toutefois, en interviewant les entreprises, cet exercice pouvait devenir une source d'information supplémentaire sur la qualité des données de l'EERH.

Le micro-appariement des données de ces deux sources est possible grâce aux deux cases du relevé de compte PD7A(TM) qui demandent aux entreprises d'indiquer le nombre de salariés à la dernière période de paie ainsi que la rémunération brute. C'est suite à la requête de Statistique Canada que ces cases ont été ajoutées en 1993, au formulaire de relevé de compte à la source.

2.2.2 Entreprises sélectionnées

Dans un premier temps, on a procédé à l'appariement d'entreprises qui avaient été identifiées comme ayant un lien unique entre l'entreprise et le numéro de PD. De ce premier appariement de 1 759 micro-données, il s'est avéré que 44 entreprises enregistraient un écart d'au moins 200 salariés entre les deux sources de données; 17 d'entre elles indiquaient un nombre de salariés plus élevé du côté de l'EERH tandis que les 27 autres montraient une situation inverse. Dans un second temps, un autre groupe de 852 entreprises à lien multiple, lesquelles sont reliées soit à plus d'un numéro de PD ou à plus d'un établissement, n'ont pu être appariées. En regardant les différences du nombre de salariés au niveau de l'entreprise, il s'est avéré que 15 entreprises affichaient une différence d'au moins 1 000 salariés.

2.2.3 Entreprises sélectionnées pour interviews

Des contraintes de temps ont restreint le nombre d'entreprises devant être contactées. On a donc rejoint un éventail d'entreprises, le plus représentatif possible en ce qui a trait à la province, au secteur d'activité économique et au nombre de liens lorsque c'était applicable. Donc, 5 entreprises à lien simple ont été contactées en décembre 1995 et 9 entreprises à lien multiple le furent en janvier 1996. Pour cette entrevue téléphonique, il n'y avait pas de questionnaire officiel; les intervieweurs signalaient aux répondants que les données de salariés ou de rémunération brute déclarées à l'EERH différaient de celles indiquées au formulaire PD7A(TM) de Revenu Canada.

2.2.4 Raisons des écarts

Voici une liste sommaire des raisons causant les écarts entre les deux sources:

Liens défectueux/manquants ou numéro de PD partagé : problème relevant de retards dans la mise à jour de la base de données du registre central (BDRC) de Statistique Canada.

Numéros de PD rapportant des données de pensions et de bien-être : ici les données déclarées à l'EERH ne sont pas en cause. Toutefois, ces numéros de PD auraient dû être identifiés afin de ne pas les utiliser, problème se rapportant à la gestion de la BDRC.

Paies manquantes: des paies de vacances auraient dues être déclarées avec les montants de paiements spéciaux de l'EERH. Aussi des paies confidentielles n'étaient pas incluses.

Façon de répondre à Revenu Canada: des faiblesses au niveau de la réponse à Revenu Canada cause des problèmes pour l'estimation. Dans certains cas, le répondant n'est pas en cause, c'est plutôt le service bancaire qui le fait pour l'entreprise.

Changements manuels: dans un cas, les changements effectués par les analystes ont contribué à abaisser les données de l'EERH.

3.1 Mesures d'amélioration de la qualité prises suite aux deux études précédentes

3.1.2 Actions entreprises afin de réduire les écarts

- Afin de recueillir les informations manquantes soient: les payes, les allocations et les paiements spéciaux, les actions suivantes ont été entreprises soient:
 - 1) réitérer les concepts de l'EERH auprès des répondants;
 - 2) revoir ces mêmes concepts avec les intervieweurs des bureaux régionaux;
 - 3) planifier de recontacter sur une base courante les entreprises de l'EERH afin de vérifier leurs structures et habitudes de réponse.
- Il est possible de croire que l'arrivée d'un système de vérification plus efficace lors du remaniement de l'EERH a engendré une diminution de l'impact des ajustements manuels effectués en cours d'enquêtes.
- Pour ce qui est de réduire les erreurs de réponse et d'entrée des données ayant un impact majeur sur les estimations d'emploi et de rémunération, on a mis en place un mécanisme de détection qui permet de détecter ces erreurs et par la suite de les corriger avant la compilation des données.
- Aussi, afin que le registre des entreprises soit plus à jour, on a référé sur une base constante au registre des entreprises tous les problèmes d'ordre structurel trouvés.

3.2 Appariement des données mensuelles de l'enquête sur la rémunération auprès des petites entreprises avec les données du relevé de compte à la source (formulaire PD7A) de Revenu Canada

3.2.1 Contexte de l'enquête sur la rémunération des petites entreprises

L'enquête sur la rémunération des petites entreprises (l'ERPE), qui fait appel à un sous-échantillon de 7 500 unités tirés de 86 000 comptes de retenues sur la paie, a été mise sur pied en mars 1994. Ces unités rassemblent les petites entreprises de moins de 100 salariés, pour lesquelles un seul établissement a été identifié. Cette enquête, qui recueille toutes les variables de l'EERH, sert à attribuer mensuellement les caractéristiques manquantes aux données administratives (PD7A), lesquelles formaient déjà 30 % des estimations d'emploi et de rémunération.

3.2.2 Appariement des micro-données

Comme le sous-échantillon de l'ERPE est tiré à partir des comptes mensuels PD7-A, appairer les micro-données fournies à Revenu Canada avec celles envoyées à Statistique Canada était donc facilement réalisable. Comme la relation entreprise, établissement et numéro de PD était unique, cette tâche était d'autant plus facile. De plus, comme les données des deux sources d'information sont disponibles mensuellement, il fut donc décidé d'effectuer l'appariement des micro-données à cette fréquence. L'établissement d'une série de données chronologiques portant sur les écarts des données appariées offrait une perspective d'analyse intéressante par rapport à la qualité des données d'enquête. Cette série chronologique fut donc établie dès le début de l'ERPE, soit en mars 1994.

Les variables appariées, comme pour les deux expériences précédentes, furent le nombre de salariés et la rémunération brute mensuelle. Des 2 500 entreprises enquêtées entre mars 1994 et avril 1995, 2 300 ont pu en moyenne être appariées. Le non-appariement de certaines entreprises est causé par des données manquantes, des erreurs majeures d'entrée de données dans l'une des deux sources ou par des raisons telles: les refus, la non-disponibilité des données du côté de l'enquête. La distribution moyenne des données appariées pour cette période est la suivante: 68 % ont enregistré un nombre identique de salariés, 16 % avaient été imputées du côté des données administratives et finalement 16 % enregistraient des écarts que l'on peut qualifier de réels.

3.2.3 Résultats des appariements

Pour la moyenne des 391 entreprises (6 %) pour lesquelles des écarts ont été enregistrés, voici quelques statistiques permettant de quantifier les écarts observés de mars 1994 à avril 1995:

Variable	Écart moyen %	Écart maximal %	Écart minimal %
Salariés	7,7	13,9 (janvier 1995)	1,0 (avril 1996)
Rémunération Brute	6,7	14,8 (avril 1995)	0,2 (février 1996)

Les écarts observés sont positifs dans 92 % du temps pour la variable salarié et dans 85 % pour la variable rémunération brute. Un écart positif signifie que les données de la source administrative montrent des observations supérieures à celles provenant de l'ERPE.

3.2.4 Raisons causant les écarts

Après analyse des écarts selon le code de réponse de l'ERPE, on a pu noter qu'il y avait un nombre important d'entreprises pour lesquelles l'écart entre les deux sources était de 100 % soit, quand celles-ci étaient fermées (code 8) et ou n'avaient rien à déclarer (code 9). Les différences observées entre les deux sources pour ces deux codes de non-réponse contribuent à accentuer la sous-évaluation des données d'enquête tant pour le nombre de salariés que pour la valeur de la rémunération brute. De ce fait, il fallait donc s'attarder plus spécifiquement aux entreprises rapportant de tels résultats. L'outil utilisé fut la liste des commentaires écrits par les intervieweurs, lesquels avaient été rendus obligatoires vu le nombre d'entreprises pour lesquelles des codes de réponse 8 et 9 avaient été employés. L'examen des commentaires se rapportant à 57 entreprises sur 74 pour le mois de septembre 1994 a révélé les choses suivantes:

- 1) lorsque le code de non-réponse mentionne que l'entreprise est fermée (code 8)
 - la date de fermeture des entreprises était inscrite dans la plupart des commentaires examinés
 - les commentaires ne sont présents en moyenne que 5 fois sur 8.
- 2) lorsque le code de non-réponse mentionne que l'entreprise n'a rien à déclarer (code 9)
 - on n'incluait pas souvent les propriétaires actifs, les associés et les autres dirigeants d'entreprises constituées en société, lesquels font partie de la définition de salarié. Les intervieweurs inscrivaient dans les commentaires qu'il n'y avait pas de salariés seulement que des propriétaires actifs.

Un autre aspect de l'appariement méritait aussi d'être examiné, soit lorsque le nombre d'employés est égal entre les deux sources mais qui toutefois montre un montant de rémunération brute différent. Examiner cet aspect amène à découvrir d'autres causes pouvant provoquer la sous-évaluation des données de l'ERPE. Une analyse de données a donc été enclenchée avec 37 des 95 entreprises qui montraient une telle situation, pour le mois de référence de novembre 1994. L'outil de travail utilisé ici fut les informations complètes du questionnaire de l'ERPE. En examinant ces informations, on a pu identifier des raisons potentielles pouvant causer des écarts, ce sont:

- 1) lorsque les paiements spéciaux sont présents, il se peut que l'écart soit dû à des périodes de paie qui peuvent différer surtout lorsque les employés sont payés soit à la semaine ou aux deux semaines et ce même si les dates des périodes de paies doivent être identiques entre les deux sources. Ceci a pour effet de causer de la surévaluation des données de l'ERPE ce qui toutefois n'est pas fréquent;
- 2) l'absence fréquente de paiements spéciaux est certainement une source majeure de sous-évaluation des données de l'ERPE tout comme elle l'était pour l'EERH.

3.2.5 Recommandations faites aux intervieweurs

Suite à ces deux analyses, il s'avérait nécessaire de clarifier le concept de propriétaire actif qui semblait ne pas avoir été bien compris par les intervieweurs. Aussi, il fallait leur expliquer l'importance d'obtenir des répondants les paiements spéciaux, lesquels croient souvent que de rapporter les paies régulières de leur entreprise est bien suffisant. De plus, pour poursuivre nos études sur la qualité des données, il était indispensable que les intervieweurs inscrivent toujours des commentaires lorsque le code de réponse est soit 8 ou 9. La division des opérations d'enquêtes s'est donc occupée de communiquer les résultats de nos recherches aux intervieweurs lesquels se retrouvent dans quatre bureaux régionaux.

3.2.6 Indicateurs de qualité

Clarifier certaines situations avec les intervieweurs ainsi que renforcer la formation est essentiel afin de maintenir et d'améliorer la qualité des données d'enquête. De plus, l'établissement et la diffusion de mesures de qualité sur une base mensuelle peut aussi contribuer à accroître en bout de ligne la qualité des estimations d'emploi et de rémunération. Présenter aux intervieweurs simplement l'écart existant entre les deux sources de données ne donnerait pas une juste évaluation de la qualité d'une certaine partie du travail accompli. Toutefois, en éliminant de la comparaison les enregistrements où le contact entre le répondant et l'intervieweur n'est pas remis en question (erreurs de saisie mineures perceptibles, erreurs volontaires ou non volontaires du répondant, différences de concepts), il est possible d'utiliser cette comparaison ajustée comme indicateur de qualité. L'évaluation de la qualité du contact entre l'intervieweur et le répondant est faite au bureau central en examinant les commentaires des intervieweurs, lesquels sont maintenant très complets. À partir du mois de référence, soit mai 1995, deux indicateurs de qualité l'un pour la variable salarié et l'autre pour la variable rémunération brute ont donc été introduit au processus de production mensuel.

4. CONCLUSION

L'appariement de micro-données a été pour la section de la coordination des enquêtes et de l'amélioration de la qualité un outil qui a permis d'une part de mieux cerner les erreurs non dues à l'échantillonnage et par la suite de mettre en place des mesures correctives. Les trois études ont clairement confirmé que les erreurs non dues à l'échantillonnage avaient tendance à sous-évaluer le niveau des estimations d'emploi et de rémunération. Ces études ont contribué à la mise sur pied d'un projet majeur d'amélioration de la qualité des données pour les entreprises complexes et la création d'un indicateur de qualité pour les intervieweurs de l'ERPE.

Comme la tendance est au remplacement d'un nombre croissant d'unités enquêtées par des données administratives, auxquelles il faut imputer des caractéristiques tirées de ces enquêtes, l'amélioration de la qualité des données d'enquêtes demeurera toujours un sujet d'actualité. Conséquemment, tant que les erreurs non dues à l'échantillonnage existeront, la section de la coordination des enquêtes et de l'amélioration de la qualité travaillera à leur élimination.

5. BIBLIOGRAPHIE

Paillé, B. *PD7A/SEPH Micro Match Analysis: An Evaluation of Administrative Employment Data*, division du travail, Statistique Canada, février 1994.

Sources et définitions

Enquête sur l'emploi, la rémunération et les heures: les résultats de cette enquête sont publiés par Statistique Canada sous le nom de 'Emploi, gains et durée du travail' au catalogue 72-002 mensuel.

Formulaires T-4, PD7A, PD7A(TM) de Revenu Canada: pour plus d'information consultez le 'Guide de l'employeur - Retenues sur la paie' renseignements de base.

Salariés: toute personne rémunérée pour les services rendus au Canada ou pour un congé payé et pour laquelle l'employeur est tenu de remplir une déclaration T-4 supplémentaire de Revenu Canada. Le terme salarié englobe les salariés à temps plein, à temps partiel et occasionnels ou temporaires.

Rémunération brute: il s'agit du montant total avant les retenues de toute rémunération versée aux salariés au cours du mois de référence. Le montant devrait correspondre au total de tous les montants qui apparaîtraient dans la case 14 du feuillet T-4. Ce montant comprend: les salaires réguliers; les commissions; les paiements des heures supplémentaires; les congés payés; les salaires à la pièce; les paiements spéciaux; et les allocations et les avantages fédéraux imposables reconnus par Revenu Canada.

Entreprise: toute entreprise commerciale ou institution, constituée ou non en corporation; il peut s'agir d'une entreprise à propriétaire unique, d'une association, d'une société ou d'une autre forme d'organisation.

Établissement: à des fins statistiques, l'établissement est la plus petite unité pouvant fournir des données sur la production économique. En général, une usine, une mine, un magasin ou à une unité semblable.

VALIDITÉ DE L'AUTO-DÉCLARATION DES PROBLÈMES DE SANTÉ CHRONIQUES LORS DE L'ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION

Gary Catlin, Karen Roberts¹ et Susan Ingram

La validité des déclarations faites durant une enquête est un élément déterminant de la confiance qu'ont les chercheurs dans les données. L'Enquête nationale sur la santé de la population (ENSP) est une enquête longitudinale effectuée tous les deux ans pour obtenir des données sur l'état de santé, sur les déterminants de ce dernier et sur l'utilisation des services de santé. En raison de la nature longitudinale de l'enquête, on accorde une importance particulière à l'exactitude des variations des caractéristiques observées au cours du temps. L'un des renseignements importants concernant l'état de santé collectés dans le cadre de l'ENSP a trait à l'existence de problèmes de santé chroniques, tels que l'asthme, l'arthrite, l'hypertension, le diabète, la migraine, l'épilepsie, les ulcères et les séquelles d'un accident vasculaire cérébral. Ces renseignements ont été collectés en 1994-1995 et sont de nouveau collectés en 1996 auprès des mêmes répondants. La présente étude sera effectuée auprès des répondants des deux premières périodes de collecte de données de l'enquête de 1996, soit un échantillon d'environ 7 000 ménages. Durant le deuxième cycle de l'enquête, nous poserons des questions pour découvrir les raisons des écarts observés entre deux points dans le temps. Les raisons des variations incluront les changements réels dus à l'apparition ou à la disparition de troubles, ainsi que les écarts attribuables à l'erreur non due à l'échantillonnage. La comparaison permettra de déterminer quels sont les types de troubles chroniques qui sont déclarés avec le plus d'exactitude et les facteurs qui ont une incidence sur l'exactitude des renseignements déclarés.

¹Karen Roberts, Division des statistiques de la santé, Statistique Canada, Parc Tunney, Ottawa, Ontario, K1A 0T6.

ENQUÊTE SUR LE TABAGISME AU CANADA

Lecily Hunter¹

RÉSUMÉ

L'enquête sur le tabagisme au Canada qui s'est déroulée entre avril 1994 et février 1995 était une enquête longitudinale comportant quatre prises de contact à intervalles de trois mois. Il s'agissait de mesurer le changement des taux de tabagisme à la suite d'une réduction des taxes sur les cigarettes. Contrairement aux attentes, l'enquête a indiqué une diminution de la prévalence du tabagisme. On a mis en doute le bien-fondé des résultats, en se demandant s'ils étaient influencés outre mesure par des erreurs non dues à l'échantillonnage.

MOTS CLÉS: Tabagisme; biais de non-réponse; libellé des questions; enquête longitudinale.

1. INTRODUCTION

1.1 Contexte

Les enquêtes menées sur le tabagisme au cours des 10 dernières années ont indiqué de façon constante des taux de tabagisme situés entre 29 et 34 % environ. Ces enquêtes ont mis en oeuvre un choix de méthodes allant des suppléments de l'enquête sur la population active jusqu'au système de composition aléatoire des appels, permettant ou non les réponses par procuration; la taille de l'échantillon variant entre 10 000 et 20 000 et la collecte se faisant en un mois ou sur une base trimestrielle ou mensuelle. Les questions avaient évolué depuis les premières enquêtes de Statistique Canada menées au cours des années 1970, à tel point que les estimations de la prévalence du tabagisme semblaient relativement fiables aux spécialistes de la santé. Néanmoins, au début des années 1990, on a remarqué une diminution importante du nombre de cigarettes fabriquées en vue du marché national, alors qu'il n'y avait pas de baisse correspondante du nombre de personnes qui fumaient ni du nombre de cigarettes qu'elles fumaient. Cependant, il y avait un écart croissant entre le coût d'achat des cigarettes au Canada et leur prix aux É.-U., surtout à cause du taux d'imposition plus élevé au Canada. On a conclu en fin de compte que la majorité des cigarettes fabriquées au Canada en vue du marché américain (qui avait connu une forte hausse au cours de la même période) étaient réimportées en contrebande et vendues sur le marché clandestin à des prix beaucoup moins élevés que le prix exigé chez le dépanneur.

En février 1994, le gouvernement fédéral a décidé de réduire les taxes sur les cigarettes canadiennes afin de lutter contre le commerce en contrebande. De nombreuses provinces ont également décidé de réduire la taxe provinciale sur les cigarettes, entraînant une diminution nette du prix d'achat au détail pouvant atteindre 50 %. Santé Canada et d'autres organismes de santé s'inquiétaient que les prix moins élevés susciteraient une prévalence plus élevée du tabagisme et une augmentation du nombre de cigarettes que l'on fumait. Par conséquent, Santé Canada a voulu mener une enquête qui permettrait de mesurer et de surveiller les changements du taux de tabagisme au cours des 12 mois qui ont suivi le changement du prix.

On s'attendait que chaque point de pourcentage dans l'estimation de la prévalence soit contesté par ceux qui appuyaient l'un ou l'autre côté de cette question controversée. On a donc choisi pour cette enquête les questions qui avaient été utilisées avec succès par d'autres enquêtes et qui dans le passé permettraient de mesurer les taux de tabagisme le plus exactement possible. Néanmoins, il a été reconnu par Santé Canada aussi bien que par Statistique Canada que certains types de comportement ne se laissent pas mesurer aussi exactement qu'on le

¹Lecily Hunter, Division des enquêtes spéciales, Statistiques Canada, 5D5 Édifice Jean Talon, Parc Tunney, Canada K1A 0T6.

voudrait à cause de facteurs comme la rapidité du changement, les états de transition (qui influencent l'aptitude des personnes à donner une réponse exacte) et la perception de l'acceptabilité sociale (qui influence le souci des personnes à donner une réponse exacte).

Statistique Canada a mené l'enquête sur le tabagisme au Canada (SoSiC) sous forme d'enquête longitudinale comportant quatre entrevues à intervalles de trois mois. Les résultats initiaux ont montré un taux de prévalence de 30 % environ, soit le même niveau qu'en 1991. Les entrevues de suivi ont indiqué des taux à la baisse, tombés à 27 % lors du quatrième trimestre. Ce résultat allait à l'encontre des attentes des spécialistes de la santé, qui ont mis en doute la validité des résultats de l'enquête. Le présent exposé fait état des différentes méthodes qui ont permis d'examiner l'information disponible afin d'évaluer l'effet possible d'erreurs non dues à l'échantillonnage.

1.2 Conception de l'enquête

L'échantillon initial a été choisi à l'aide de la méthode de composition aléatoire des appels (Random Digit Dialling). On a choisi dans chaque ménage un répondant âgé de 15 ans ou plus. Ce choix s'est fait selon une probabilité inégale afin de satisfaire l'exigence de fiabilité de Santé Canada pour chacun des quatre groupes d'âge : 15-19, 20-24, 25-64 et 65+. De plus, on a éliminé un certain nombre de ménages comportant uniquement des personnes de 25-64 ans afin de réduire les coûts.

On a communiqué de nouveau avec les répondants de cette interview initiale à des intervalles de trois, six et neuf mois afin de satisfaire la deuxième exigence de Santé Canada, c'est-à-dire de pouvoir détecter des changements relativement faibles du taux de prévalence du tabagisme.

Les entrevues se sont déroulées en utilisant le mode ITAO (Interviews téléphoniques assistées par ordinateur), qui a permis de satisfaire la troisième exigence de Santé Canada, c'est-à-dire d'aller sur le terrain et de recueillir les données rapidement. Un ensemble de questions de base a été utilisé à chaque trimestre afin de déterminer la situation courante des répondants relativement au tabagisme, mais le reste des questions a varié à chaque trimestre afin de permettre l'étude de différents aspects du tabagisme. La collecte des données a eu lieu en avril-mai, en août et novembre 1994 et en février 1995.

1.3 Les attentes

De nombreux spécialistes de la santé s'attendaient à observer une augmentation appréciable tant de la proportion de personnes qui fumaient que du nombre de cigarettes qu'elles fumaient. Cette attente se fondait sur 1) les données sur la fabrication des cigarettes et 2) des calculs fondés sur l'«élasticité par rapport au prix» (des prix moins élevés entraînant une demande accrue).

Les jeunes de moins de 20 ans étaient considérés comme le groupe le plus vulnérable. Les spécialistes estimaient que ce groupe était plus attentif au prix et qu'il s'adonnerait au tabagisme en plus grand nombre, garderait ce comportement plus longtemps et fumerait un plus grand nombre de cigarettes chaque jour. Ce phénomène neutraliserait tout progrès accompli dans la sensibilisation des jeunes aux dangers du tabagisme en vue de les empêcher de commencer à fumer.

1.4 Les résultats

Tableau 1. Estimations du taux de tabagisme et du nombre de cigarettes que l'on fumait, par cycle de collecte (SoSiC)

Cycle de collecte	N ^{bre} de répondants	Proportion de fumeurs	N ^{bre} moyen de cigarettes que l'on fume
1 (Avril-mai 1994)	15 804	Actuels- 30,4 % Quotidiens - 24,8 % Occasionnels - 5,7 %	Actuellement - 15,8 Quotidiennement - 18,7 Occasionnellement - 3,2
2 (Août 1994)	13 398 (85 % de la Q1)	Actuels- 29,9 % Quotidiens - 24,0 % Occasionnels - 6,0 %	Actuellement - 15,6 Quotidiennement - 18,7 Occasionnellement - 3,2
3 (Novembre 1994)	12 808 (81 % de la Q1)	Actuels - 29,5 % Quotidiens - 24,6 % Occasionnels - 4,9 %	Actuellement - 15,6 Quotidiennement - 18,3 Occasionnellement - 1,8
4 (Février 1995)	12 424 (79 % de la Q1)	Actuels - 27,4 % Quotidiens - 23,2 % Occasionnels - 4,2 %	Actuellement - 15,3 Quotidiennement - 17,7 Occasionnellement - 2,0

11 119 personnes ont répondu aux questions de l'enquête à chacun des quatre trimestres.

1.5 Réaction

Lorsque les résultats du premier trimestre ont été publiés, la réaction immédiate tant des médias que des groupes de non-fumeurs a été que l'enquête devait être inexacte puisque le chiffre de 30,4 % (fumeurs actuels) n'était pas «suffisamment élevé». En réalité, il n'y avait aucune différence manifeste relativement à l'enquête antérieure de Statistique Canada mesurant les taux de tabagisme : l'enquête sociale générale (ESG) de 1991 estimait à 31 % le nombre de fumeurs «actuels» (26 % quotidiens, 5 % occasionnels). [NOTA : l'ESG91 n'a pas calculé le nombre moyen de cigarettes que l'on fumait, mais a demandé quel était le nombre de cigarettes que l'on fumait, en fonction d'intervalles.]

À chaque trimestre subséquent, les résultats ont montré une baisse de la prévalence du tabagisme, en particulier lors du quatrième trimestre, et les critiques au sujet de la fiabilité des résultats sont devenues plus explicites. Pour s'attaquer aux problèmes, il fallait vraiment répondre à la question : Est-ce que cette enquête décrit «la réalité» ou les résultats sont-ils contaminés par une erreur non imputable à l'échantillonnage?

2. ÉVALUATION DES RÉSULTATS

2.1 Non-réponse à la première entrevue

Parmi les 29 149 cas identifiés comme des ménages dans l'échantillon, 25 598 ont donné une réponse au niveau des ménages. Cela signifie que nous avons pu recueillir la liste des ménages et choisir une personne comme répondant ou éliminer le ménage.

Afin de déterminer si notre échantillon initial de 25 598 ménages répondants était représentatif de la population, nous avons comparé la composition des ménages avec les données de l'enquête sur la population active (EPA) de mai 1994 et avec les données de l'ESG de 1994 (tableau 2).

Tableau 2. Proportion de ménages selon la composition des ménages : SoSiC, EPA et ESG

	SoSiC - avril-mai 1994 (non pondérée)	EPA - mai 1994 (pondérée)	ESG de 1994 (non pondérée)
Ménage d'une personne	20,3	24,3	24,2
Une personne de 15 ans ou plus et au moins une personne de moins de 15 ans	3,5	3,9	3,2
Deux personnes de 15 ans ou plus et au moins une personne de moins de 15 ans	20,1	19,3	16,9
Deux personnes de 15 ans ou plus et sans enfant de moins de 15 ans	29,7	30,6	32,0
Plus de deux personnes de 15 ans ou plus avec ou sans enfant de moins de 15 ans	26,3	22,0	23,7

Les chiffres du tableau 2 semblent indiquer que la SoSiC comportait une certaine sous-représentation des ménages moins nombreux (une personne) et une sur-représentation des ménages plus nombreux. Il n'y a là rien d'inusité pour une enquête RDD puisqu'il est normalement plus facile de joindre quelqu'un à la maison lorsque le ménage compte un plus grand nombre de personnes. Deux facteurs ont probablement influencé les résultats : la durée de la période de collecte et le système de planification des appels.

Même s'il est plus difficile de joindre quelqu'un à la maison dans un ménage peu nombreux, le fait de prolonger la période de collecte d'une ou deux semaines suffit à réduire appréciablement les cas de non-contact (la principale source de non-réponse de la SoSiC). La période de collecte de cette enquête a été de six semaines, ce qui semble avoir été une période trop courte pour la taille de l'échantillon.

Le système automatisé de planification des appels a également eu un effet sur la collecte. Une bonne partie de l'échantillon n'a reçu son premier appel qu'après la première moitié au moins du cycle de collecte, bien que certains cas aient donné lieu à plusieurs appels sans réponse. Certains systèmes de planification des appels plus sophistiqués permettent de classer les cas en fonction de l'évaluation d'un certain nombre de critères; les critères utilisés et l'importance qui leur est accordée sont des facteurs flexibles et peuvent être fixés par l'équipe de conception de l'enquête. À cause du caractère plus «primitif» du système de planification des appels utilisé pour l'enquête SoSiC, de nombreux cas offraient une fenêtre plutôt étroite pour établir un contact.

2.2 Non-réponse lors des interviews de suivi

La première entrevue de la SoSiC comptait 15 804 répondants. Même si l'on suppose que l'échantillon initial était représentatif, il est possible que la non-réponse aux interviews de suivi ait biaisé les résultats des trimestres subséquents. Afin de vérifier cette hypothèse, nous avons d'abord supposé que les personnes qui ont répondu étaient franches et exactes; pour ce qui est des personnes qui n'ont pas répondu, nous avons supposé qu'elles fumaient durant le ou les trimestres manquants. Différentes stratégies d'imputation ont été utilisées et les estimations résultantes ont été comparées (tableau 3).

Tableau 3. Taux de tabagisme, après imputation en cas de non-réponse aux interviews subséquentes, selon diverses stratégies d'imputation

	Interview 1	Interview 2	Interview 3	Interview 4
Pas d'imputation	30,4	29,9	29,5	27,4
1. On attribue aux non-répondants leur dernier comportement connu de fumeur/non-fumeur	30,4	30,3	30,1	28,4
2. On attribue aux non-répondants le comportement fumeur/non-fumeur établi à la première interview	30,4	30,3	30,0	28,5
3. Comme en 1, mais si la personne a cessé de fumer entre janvier et mai 1994, elle est considérée comme fumeur	30,4	30,4	30,3	28,8
4. Comme en 1, mais si la personne a changé de comportement durant la période de l'enquête, elle est considérée comme fumeur	30,4	30,4	30,6	29,0
5. Comme en 3, mais tous les non-répondants de moins de 20 ans sont considérés comme fumeurs	30,4	31,1	31,3	29,8

Même en déterminant de façon assez radicale le comportement des non-répondants par rapport au tabagisme, dans tous les cas les estimations indiquent une baisse des taux de tabagisme entre le troisième et le quatrième trimestre. Il semble donc que l'effet observé n'a pas été causé par des erreurs de non-réponse, mais reflète ce que les gens ont réellement signalé comme changement de comportement. Il reste donc une autre question au sujet des erreurs non dues à l'échantillonnage : est-ce que l'enquête a mesuré des changements réels ou est-ce que les répondants ont signalé leur comportement de façon incorrecte?

2.3 Courbe d'apprentissage (familiarité avec les questions lors d'interviews subséquentes)

Un des effets éventuellement négatifs des enquêtes longitudinales est que les répondants se familiarisent avec les questions et savent quelles réponses mettront fin à l'interview plus rapidement. On s'attendrait avec le temps que les gens suivent la «voie la plus courte» si tel était le cas.

Pour ce qui est de la SoSiC, chaque entrevue a duré dix minutes environ. Les fumeurs et les non-fumeurs ont répondu plus ou moins au même nombre de questions; l'interview n'a guère duré qu'une minute de plus pour les fumeurs. Les répondants qui signalaient un changement de comportement depuis l'entrevue antérieure ont dû répondre à deux autres questions afin de déterminer quand et pourquoi le changement avait eu lieu.

Dans l'ensemble, 89,2 % des répondants n'ont signalé aucun changement de comportement au cours des quatre interviews. Parmi les autres (11,8 %), un peu plus de la moitié ont signalé un changement tandis que le reste a signalé deux changements ou plus. Le tableau 4 indique la proportion de répondants (ayant répondu à chacun des quatre trimestres) qui a signalé un changement de comportement (dans un sens ou dans l'autre). Il ne semble pas que ces résultats montrent une tendance appuyant l'hypothèse d'un apprentissage. Toutefois, on peut tirer une conclusion importante des données sur le changement de comportement. L'estimation de changement à partir de données recueillies rétrospectivement représente la moitié environ de l'estimation mesurée directement; il en est ainsi probablement parce que les transitions comme le fait de cesser de fumer ont tendance à être fréquentes et de courte durée, de sorte qu'elles sont oubliées ou laissées de côté lors d'une question rétrospective.

Tableau 4. Proportion de répondants longitudinaux signalant un changement de comportement au cours de la période à l'étude

	ont changé de comportement (%)	se sont mis à fumer (%)	ont cessé de fumer (%)
Janv. 94/avr.-mai 94 (question rétrospective)	1,6	0,6	1,0
Avr.-mai 94/août 94	4,9	2,1	2,8
Août 94/nov. 94	4,3	2,1	2,3
Nov. 94/févr. 95	4,1	1,1	3,0

2.4 Sensibilité/acceptabilité sociales

On a suggéré que les campagnes contre le tabagisme menées au cours des dix dernières années ont attaché une certaine honte sociale au tabagisme. Si tel est le cas, il est possible que les répondants ne veuillent pas admettre qu'ils fument. Il y a peu de données pour appuyer cette proposition. Près du quart des répondants de la SoSiC ont signalé qu'ils fumaient au cours des quatre interviews. Au cours des dix dernières années, les taux de tabagisme ont changé relativement peu dans l'ensemble, mais il y a eu un déplacement majeur entre l'usage quotidien et occasionnel du tabac. Si l'acceptabilité sociale joue un rôle dans la façon dont les gens répondent aux questions sur le tabagisme, cette influence s'exerce fort probablement sur la façon de signaler la fréquence de l'usage du tabac ou le nombre de cigarettes que l'on fume, et non pas sur la tendance à nier complètement que l'on fume.

2.5 Libellé des questions

Le libellé des questions et le contexte dans lequel on les pose exerce une influence sur la qualité des réponses. Puisque plusieurs enquêtes ont servi à recueillir des renseignements sur le tabagisme à peu près à la même période, nous avons pu comparer les résultats et leurs liens avec le libellé (tableau 5).

Le fait de poser d'abord des questions sur le comportement actuel (ESG91, SoSiC et ENSP) semble produire une estimation très différente des fumeurs actuels relativement à la stratégie de poser d'abord des questions sur le comportement antérieur (CADS et ESG95). Les différences se rapportent surtout au sous-groupe des fumeurs «occasionnels». Ce groupe est devenu plus nombreux relativement aux estimations d'il y a une dizaine d'années, tant à cause de l'accroissement du nombre de personnes qui tentent de cesser de fumer ou qui commencent à fumer qu'à cause du nombre accru de restrictions quant aux endroits où l'on peut fumer.

Toutes ces enquêtes adoptent comme stratégie de demander au répondant de se classer comme fumeur ou non-fumeur, les questions subséquentes étant déterminées en fonction de cette classification. Pour la plupart des gens, cela ne pose pas de problème réel, mais pour les 5-10 % de la population qui changent de comportement une fois ou plus au cours d'une période de 12 mois, ce choix n'est pas nécessairement clair. Pour ce qui est de CADS et de l'ESG95, la question «avez-vous déjà été fumeur?/avez-vous déjà fumé des cigarettes?» risque d'être mal interprétée par les personnes pour qui l'usage du tabac n'est pas un comportement régulier. Si le répondant fume à l'occasion ou ne fait que commencer à fumer, il aura peut-être tendance à donner une réponse négative, ce qui l'exclut de toutes les questions subséquentes au sujet du tabagisme.

Pour les enquêtes à venir, une meilleure stratégie serait peut-être d'établir une série de questions qui mesure des comportements particuliers au cours de périodes précises. À l'étape analytique, nous établissons dès lors nos définitions en fonction de ces comportements au lieu de demander aux répondants de se classer eux-mêmes et de laisser cette classification orienter les questions subséquentes.

Tableau 5. Libellé des questions et estimation résultante du taux de tabagisme d'après des enquêtes récentes sur le tabagisme

Enquête	Libellé des questions	Taux de tabagisme
ESG91² (collecte mensuelle - estimation annuelle pour 1991)	Les prochaines questions se rapportent à l'usage des cigarettes. Fumez-vous des cigarettes chaque jour, occasionnellement ou pas du tout?	31 % actuellement 26 % quotidiennement 5 % occasionnellement
SoSiC³ (première entrevue en avril- mai 1994)	(Introduction) À l'heure actuelle fumez-vous des cigarettes chaque jour, occasionnellement ou pas du tout?	30 % actuellement 25 % quotidiennement 6 % occasionnellement
ENSP⁴ (collecte trimestrielle - estimation annuelle pour avril 1994-juin 1995)	Les quelques questions suivantes se rapportent au tabagisme. 1. Est-ce que quelqu'un dans ce ménage fume régulièrement à la maison? 2. À l'heure actuelle est-ce que... fume des cigarettes chaque jour, occasionnellement ou pas du tout?	31 % actuellement 26 % quotidiennement 5 % occasionnellement
CADS⁵ (collecte en octobre 1994)	Je vais maintenant vous poser des questions au sujet du tabagisme. 1. Avez-vous déjà fumé des cigarettes? 2. À quel âge avez-vous commencé à fumer? 3. Fumez-vous actuellement des cigarettes? 4. Au cours des 7 derniers jours, avez-vous fumé des cigarettes? 5. Avez-vous fumé le même nombre de cigarettes chaque jour?	27 % actuellement (pas de distinction entre un usage quotidien et occasionnel)
ESG95⁶ (collecte mensuelle - estimation annuelle pour 1995)	Je vais maintenant vous poser quelques questions sur votre exposition à la fumée des cigarettes. 1. Avez-vous déjà fumé des cigarettes? 2. Fumez-vous actuellement des cigarettes? 3. Fumez-vous normalement des cigarettes chaque jour?	26 % actuellement 25 % quotidiennement 2 % occasionnellement

² Enquête sociale générale (janvier - décembre 1991)

³ Enquête sur le tabagisme au Canada, cycle 1 (avril-mai 1994)

⁴ Enquête nationale sur la santé de la population (avril 1994 - juin 1995)

⁵ Enquête canadienne sur la consommation d'alcool et d'autres drogues (octobre 1994)

⁶ Enquête sociale générale (janvier - décembre 1995)

2.6 Effets externes

Le tabagisme est un phénomène volatil qui est donc difficile de mesurer. Même si le taux de tabagisme peut sembler stable au cours d'un certain temps, le comportement des personnes mêmes qui fument ou ne fument pas change rapidement. Au moment de la collecte du quatrième trimestre de la SoSiC (février 1995), le gouvernement fédéral venait tout juste d'entreprendre une série d'annonces télévisées informant le grand public des risques associés à la fumée secondaire chez les bébés et les jeunes enfants. Plusieurs administrations provinciales avaient récemment mené d'autres campagnes anti-tabac; l'Ontario venait de mettre en oeuvre des lois interdisant la vente de cigarettes dans les pharmacies et obligeant les magasins à afficher des annonces avertissant les mineurs qu'ils ne pourraient pas acheter de cigarettes sans pièce d'identité; et le gouvernement fédéral avait mis en oeuvre une faible augmentation de la taxe sur les cigarettes. Il est possible que certains de ces facteurs aient pu pousser les gens (au moins provisoirement) à cesser de fumer.

Les fluctuations saisonnières ont peut-être également un rôle à jouer dans l'usage du tabac. Certaines personnes cessent de fumer lorsqu'elles pratiquent un sport; d'autres décident de cesser de fumer comme résolution du nouvel an; les jeunes commencent parfois à fumer à l'approche de la rentrée scolaire. Nous n'avons pas de données nous permettant de mesurer ces facteurs de façon fiable, mais certains travailleurs de la santé estiment qu'une telle influence est possible.

Pour ce qui est des données SoSiC, les changements de comportement les plus importants se sont produits entre novembre 1994 et février 1995 lorsque 1,1 % environ de la population est passé de la catégorie non-fumeur à fumeur et que 3 % environ de la population est passé de la catégorie fumeur à non-fumeur, soit une baisse nette de 2 % du taux de tabagisme. À peu près 40 % des gens qui «ont cessé» de fumer avaient été des non-fumeurs lors de l'une des interviews antérieures, tandis que les autres 60 % avaient été des fumeurs à chacune des trois interviews précédentes. Compte tenu des raisons fournies par les gens pour leur changement de comportement, il n'y avait aucune différence apparente entre le quatrième trimestre et les trois trimestres antérieurs : la plupart des gens ont affirmé avoir cessé de fumer surtout pour des raisons de santé ou à cause du coût du tabac.

2.7 Comment mesurer une cible en mouvement

Un des principaux obstacles à la mesure de la prévalence du tabagisme est son changement perpétuel. Il y a toujours une partie de la population qui se trouve en état de transition (qui commence à fumer ou cesse de fumer) à un moment donné. Il est difficile pour ces gens de répondre aux questions de façon rationnelle puisque les enquêtes sont normalement conçues en fonction de catégories distinctes que l'on peut analyser (catégories fumeur ou non-fumeur par exemple). Une solution partielle serait que les concepteurs reconnaissent l'existence de ces groupes de transition et adaptent les questions et les définitions en conséquence.

Dans le tableau 6, il y a stabilité pour 74 % environ de la population, c'est-à-dire les personnes de 25 ans ou plus qui n'ont jamais fumé, ou qui fument régulièrement sans indication qu'ils tenteront de cesser, ou encore qui ont déjà fumé mais qui ne fument pas depuis 5 ans ou plus. Les autres 26 % de la population représentent le groupe qui «risque» de changer de comportement pour ce qui est de l'usage du tabac : (4) les personnes de 25 ans ou plus qui ont cessé de fumer puis ont recommencé; (5, 6, 7) toutes les personnes de 15 à 25 ans, qui risquent le plus de commencer à fumer, même si plusieurs d'entre elles ne fumeront pas longtemps. Or même parmi les personnes à risque élevé, une partie seulement se trouvera en transition à un moment donné, de sorte qu'une estimation exacte de la prévalence du tabagisme dépend réellement de notre aptitude à mesurer un segment relativement faible mais influent de la population. Suivant le comportement de ces personnes au moment de l'enquête, nous pourrions obtenir des estimations allant de 24 % à 36 %.

Tableau 6. Répartition de la population en groupes selon le risque de changement de l'usage du tabac

25 ans ou plus, n'a jamais fumé	faible risque de changement	30 %
25 ans ou plus, fumeur au cours des 4 cycles	faible risque de cessation (au moins dans un avenir rapproché)	20 %
25 ans ou plus, a déjà fumé il y a 5 ans ou plus	faible risque de changement	20 %
25 ans ou plus, a récemment cessé de fumer (il y a moins de 5 ans) ou a fumé puis cessé de fumer au cours des 4 cycles	risque élevé de changement (comme fumeur et non-fumeur)	13 %
15-24 ans, fumeur au cours des 4 cycles	faible risque de cessation (au moins dans un avenir rapproché)	4 %
15-24 ans, n'a jamais fumé	risque élevé (de commencer à fumer)	10 %
15-24 ans, a fumé puis cessé de fumer, ou a essayé de fumer dans le passé	risque élevé de changement (comme fumeur et non-fumeur)	4 %

3. CONCLUSIONS

3.1 Qu'est-il arrivé en réalité aux taux de tabagisme?

Il s'agit vraiment de répondre à deux questions : (1) Pourquoi les taux de tabagisme n'ont-ils pas augmenté après la baisse des taxes sur les cigarettes (et donc du prix du marché)?, (2) Pourquoi la SoSiC a-t-elle indiqué une baisse des taux de tabagisme au cours de la période des quatre interviews? Une bonne partie de ce qui a été observé durant les mois qui ont suivi la réduction de février 1994 du coût des cigarettes peut s'expliquer (à tort ou à raison) en fonction de facteurs qui n'ont rien à voir avec des problèmes dus ou non dus à l'échantillonnage de l'enquête. Quelques exemples :

- La majorité des personnes pour qui le prix était un aspect important de la décision de fumer ou du nombre de cigarettes à fumer avaient déjà adapté leur comportement en fonction des prix moins élevés en contrebande, bien avant la baisse des prix «officiels».
- Des prix moins élevés pour les cigarettes libèrent une certaine quantité d'argent chaque semaine pour les personnes qui fument régulièrement, mais cela ne signifie pas que chacun utilisera cet argent pour acheter d'autres cigarettes.
- La population vieillit et cette nombreuse génération du baby-boom atteint maintenant le «temps le plus propice» pour cesser de fumer (40 ans et plus). À mesure que ces personnes cessent de fumer, c'est l'élément démographique qui entraînera une baisse des taux de tabagisme.

3.2 Quel a été l'effet des erreurs non dues à l'échantillonnage et qu'est-ce que nous avons appris?

Les erreurs non dues à l'échantillonnage ont certainement influencé les estimations de la prévalence du tabagisme, même si l'ampleur de cette influence est difficile et peut-être même impossible à mesurer. Les principales influences, que l'on pouvait prévenir, étaient les suivantes :

- (1) l'effet du système de planification des appels auprès des ménages, qui a influencé le type de ménage qu'on a pu joindre;
- (2) l'effet du libellé des questions (lié aux définitions de l'enquête), qui a rendu difficile, pour une partie des répondants, la tâche de se classer correctement;
- (3) le manque de sensibilité aux types et au nombre de transitions survenues au sein de la population enquêtée (également lié aux définitions de l'enquête).

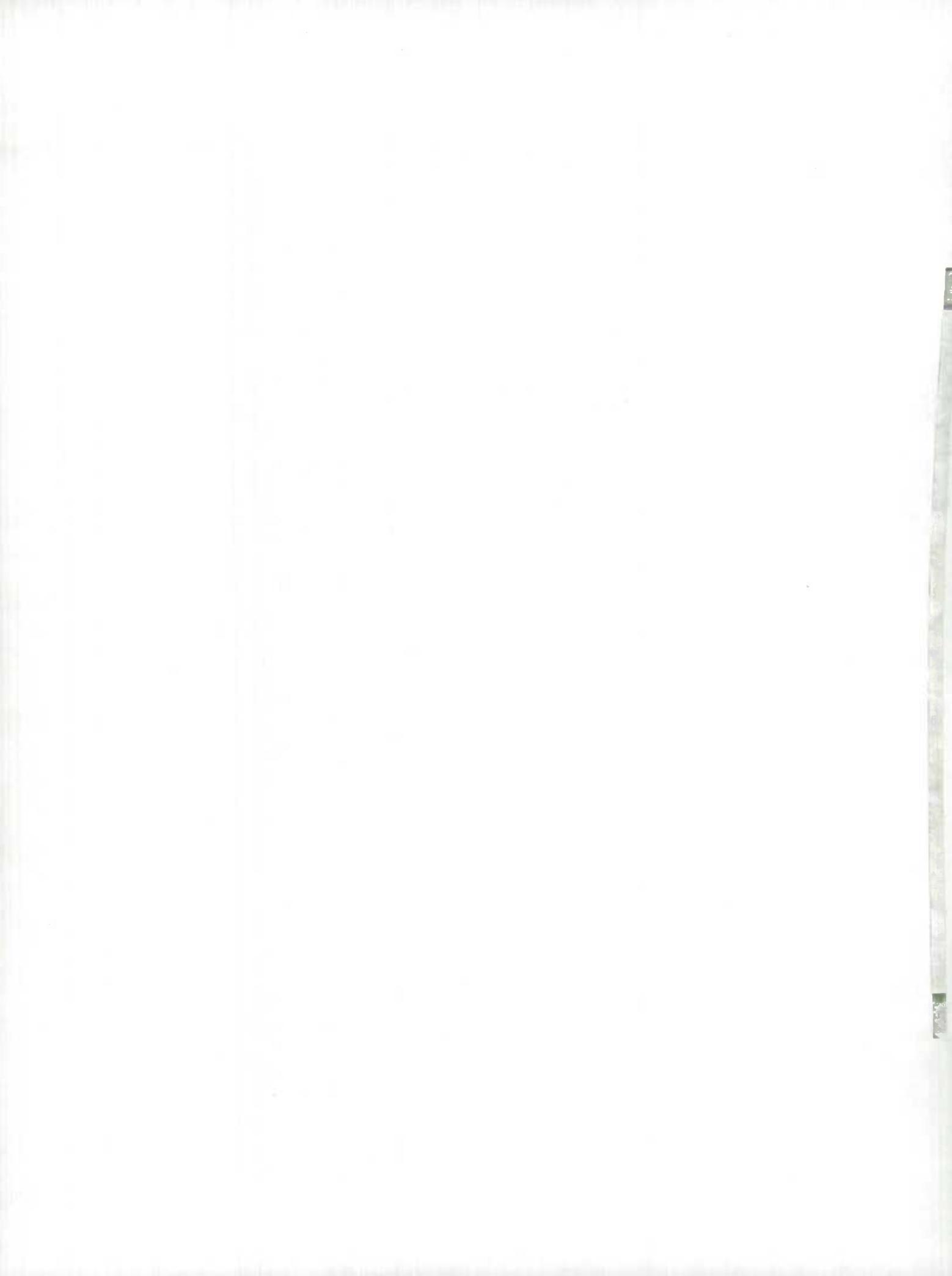
L'influence du système de planification des appels sur les estimations a probablement été minime puisque la stratification a posteriori dans l'étape de la pondération a sans doute permis de corriger une bonne partie du déséquilibre. Néanmoins, nous préparons actuellement un nouveau système de planification des appels qui permettra aux concepteurs et aux gestionnaires de mieux maîtriser la situation.

Pour ce qui est des deux autres effets, mieux vaut laisser les spécialistes de la mesure de l'usage du tabac évaluer ce que cette enquête (et d'autres enquêtes récentes) a révélé au sujet du rapport entre le besoin analytique de classer les gens de façon rationnelle et les outils nécessaires à l'atteinte de cet objectif. Il faut donc reconnaître et comprendre les comportements sous-jacents des populations que nous voulons enquêter, puis élaborer des définitions qui se fondent sur ces comportements, qui peuvent être transformées ensuite en classifications que nous avons utiliser. C'est presque le contraire de notre approche actuelle qui consiste à établir des classifications et ensuite à concevoir des questions d'enquête pour demander aux personnes à quelle catégorie elles appartiennent.

Malgré les meilleurs efforts des concepteurs d'enquête, il ne sera jamais facile de produire des estimations pour des populations qui changent constamment. Lorsque le phénomène est aussi controversé que celui-ci, avec des spécialistes de la santé, des groupes de non-fumeurs et des compagnies de tabac menant un débat sur le sens et la fiabilité de chaque point de pourcentage, aucune estimation d'enquête ne sera acceptée telle quelle. La tenue d'une enquête dans un tel contexte représente un défi pour tous, et il est essentiel que chacun comprenne les limites de nos instruments de collecte (les questions d'enquête) lorsqu'il s'agit de mesurer certains types de comportement.

SESSION 7

ERREURS DE MESURE



ERREUR DUE À L'ÉVALUATEUR LORS DE L'ÉVALUATION DU RENDEMENT DE L'INTERVIEWEUR

Paul P. Biemer¹

RÉSUMÉ

Les erreurs de l'intervieweur dans le travail d'enquête devenant un phénomène de plus en plus reconnu et préoccupant, l'évaluation du rendement de l'intervieweur est désormais un élément important des programmes d'amélioration de la qualité des enquêtes. De nombreuses méthodes permettent d'évaluer le rendement de l'intervieweur, dont la mise en place d'un programme de réinterview, la surveillance téléphonique de l'intervieweur, l'observation en personne de l'intervieweur par un superviseur ou un intervieweur principal ainsi que le codage du comportement des intervieweurs d'après l'enregistrement audio ou vidéo des interactions de l'intervieweur et du répondant. Quand l'évaluation a pour objet d'améliorer le rendement de l'intervieweur, on compare habituellement ce dernier à celui d'un autre intervieweur, à la moyenne du groupe ou à d'autres normes de rendement.

Le problème dont traite la présente communication survient quant les mesures du rendement de l'intervieweur sont biaisées et ne reflètent pas fidèlement le rendement réel, principalement à cause de l'erreur liée à l'évaluateur. Quand les critères d'évaluation du rendement sont les mêmes pour tous les intervieweurs, la principale source de l'incohérence des évaluations est l'évaluateur, autrement dit, le réintervieweur, le surveillant, le codificateur du comportement, l'observateur ou tout autre personne chargée de l'évaluation. En premier lieu, nous passons en revue les publications sur le coefficient d'objectivité en nous intéressant surtout au modèle statistique utilisé pour estimer la variance et aux conditions qui doivent être remplies pour pouvoir effectuer l'estimation, et nous illustrons son utilisation au moyen de deux exemples — l'un provenant d'une opération de surveillance téléphonique effectuée par le *Research Triangle Institute (RTI)* et l'autre tiré du *Current Population Survey Reinterview Program* du *Bureau of the Census*.

¹Paul P. Biemer, Chief Scientist, Research Triangle Institute, 304 0 Cornwallis Road, P.O. Box 12194, Research Triangle Park, North Carolina, U.S.A.

CAUSES DE LA NON-RÉPONSE À UNE QUESTION LIÉE AU TEMPS, AU RÉPONDANT ET À L'INTERVIEWEUR DANS L'ÉCHELLE D'ÉVALUATION DE L'ÉTAT DÉPRESSIF CES-D : UN MODÈLE MULTINIVEAU¹

Pieter van den Eeden², Johannes H. Smit³ et Aart-Jan Beekman⁴

RÉSUMÉ

On a déjà consacré énormément d'efforts à la réduction des effets du biais de non-réponse à une question, mais l'étude des causes de ce phénomène a moins attiré l'attention des chercheurs. Dans le présent article, nous examinons les causes liées au temps, au répondant et à l'intervieweur. Nous cherchons à construire un modèle général d'analyse de ces causes et à illustrer ce modèle. Comme on peut considérer que les cycles d'enquête sont subordonnés aux répondants, et que ces derniers sont subordonnés aux intervieweurs, le modèle d'analyse multiniveau est approprié. La non-réponse à une question est définie comme une variable dépendante de nature dichotomique, de sorte qu'il est possible d'utiliser l'analyse multiniveau de type régression logit. Les données analysées sont tirées du projet LASA (*Longitudinal Aging Study Amsterdam*) et portent sur 655 personnes âgées de 55 à 85 ans tirées d'un échantillon représentatif national des Pays-Bas; elles ont été recueillies en sept occasions séparées. L'analyse de la non-réponse à une question porte sur 20 questions de l'échelle d'évaluation de l'état dépressif du Centre d'épidémiologie (CES-D). Comme l'étude porte à la fois sur des interviews et sur des questionnaires postés à remplir soi-même, elle nous permet d'évaluer l'effet net du mode d'interview ainsi que les effets de la non-réponse à une question dans un plan d'expérience longitudinal. On observe que la non-réponse à une question est moins fréquente dans le cas des questionnaires à remplir soi-même que dans le cas des interviews, que sa fréquence diminue graduellement avec le temps, que les principaux facteurs qui déterminent la non-réponse ne sont pas le sexe ni l'âge des répondants, mais les difficultés d'ordre physique qu'ils éprouvent dans leur vie de tous les jours, la dépression et la douleur, et que la non-réponse est plus élevée pour les questions à formulation négative que pour celles à formulation positive. Aucun effet de l'interviewer n'a été décelé.

MOTS CLÉS: Modèles linéaires mixtes; composantes de la variance; recherche longitudinale; non-réponse à une question.

1. INTRODUCTION

L'expression «non-réponse à une question» désigne l'absence de données pour une question particulière posée dans le cadre d'une enquête à laquelle le répondant échantillonné avait préalablement convenu de participer. Diverses raisons peuvent expliquer la non-réponse : l'intervieweur peut négliger de poser la question; le répondant peut refuser d'y répondre ou se trouver incapable de fournir une réponse adéquate (se prêtant à une notation); l'intervieweur peut négliger de noter la réponse fournie (Groves, 1989 : p.156).

Dans la présente étude, nous nous penchons sur la fréquence et les causes de la non-réponse longitudinale dans l'échelle d'évaluation de l'état dépressif CES-D (*Center for Epidemiological Depression Scale*, Radloff, 1977). Les données ont été tirées de l'étude LASA (*Longitudinal Aging Study Amsterdam*), qui porte sur 3 107 répondants dont 662 ont été sélectionnés et incorporés dans

¹ Cette étude est fondée sur des données recueillies dans le contexte de la *Longitudinal Aging study Amsterdam* (LASA), réalisée au Département de psychiatrie de la Faculté de médecine et au Département de sociologie et de gérontologie sociale de la Faculté des sciences sociales et culturelles de la *Vrije Universiteit* d'Amsterdam. Elle a été financée par le ministère d'État à la santé, au bien-être et au sport des Pays-Bas.

² Département de méthodologie en matière de recherche sociale, *Vrije Universiteit*, De Boelelaan 1081c, 1081 HV Amsterdam, Pays-Bas. C. élec. : PJWM.vd_Eeden@scw.vu.nl.

³ Département de sociologie et de gérontologie sociale, *Vrije Universiteit*.

⁴ Département de psychiatrie, *Vrije Universiteit*.

une étude de suivi sur l'évolution de l'état dépressif. Pour l'étude de suivi, on a utilisé six questionnaires postés à remplir soi-même, expédiés à intervalles d'environ cinq mois. Cette étude se distingue des études antérieures par la composition de l'échantillon et par le mode de réponses. Il s'agit, à notre connaissance, de la première enquête de type longitudinal à examiner la non-réponse à une question et à être fondée sur une interview suivie de plusieurs questionnaires envoyés par la poste.

Les intervieweurs qui travaillent à une enquête ont habituellement chacun un groupe de répondants à questionner, ce qui fait que les répondants sont typiquement subordonnés aux intervieweurs. Dans les recherches longitudinales, les cycles d'enquête sont subordonnés aux répondants. Ainsi, le modèle d'analyse multiniveau à trois niveaux convient pour la détection des effets du temps et des caractéristiques des répondants et des intervieweurs.

Dans la section 2, nous présenterons certains concepts théoriques et certaines hypothèses concernant la non-réponse à une question. Dans la section 3, nous décrivons les données et le plan d'expérience. Dans la section 4, nous présenterons les résultats et nous examinerons brièvement la non-réponse relative à l'unité, puisqu'elle a pour effet de biaiser les résultats de l'analyse de la non-réponse à une question. Nous concluons dans la section 5 avec un résumé et une discussion des questions traitées.

2. PERSPECTIVE THÉORIQUE

Dans le présent article, nous mettons l'accent sur un certain nombre de facteurs liés à la non-réponse à une question, étudiés dans le cadre d'une enquête longitudinale. Ces facteurs sont :

Le temps. On s'est peu attardé jusqu'à maintenant à la non-réponse à une question observée dans les recherches longitudinales. Même s'il est difficile de formuler une prédiction, on peut supposer que si la non-réponse à une question est l'expression d'un manque de motivation de la part du répondant et que cette motivation diminue avec le temps, la non-réponse à une question augmentera avec le temps.

Le mode de collecte des données. On présume de façon générale que les interviews directes produisent moins de non-réponse à une question puisque l'intervieweur peut à loisir répéter ou expliquer les questions et encourager le répondant à fournir une réponse. Groves et Fultz (1985) relèvent que les interviews téléphoniques donnent typiquement un plus grand nombre de données manquantes que les interviews directes. Dans une méta-analyse des études de comparaison des modes portant sur la non-réponse à une question, De Leeuw (1992) observe en effet que les interviews directes présentent un taux légèrement inférieur de non-réponse aux questions. Cependant, elle relève que les questionnaires postés à remplir soi-même donnent de meilleurs résultats lorsque les questions abordées sont plus délicates. Ces résultats sont confirmés dans une expérience au cours de laquelle trois modes différents d'enquête sont comparés (De Leeuw, 1992).

Les caractéristiques des répondants. Les caractéristiques qu'on mentionne le plus souvent en rapport avec la non-réponse à une question plus élevée sont l'âge, le sexe, le degré de scolarité et la santé (cognitive) (Hox, De Leeuw et Kreft, 1991). Les effets dus à l'âge ont été reconnus notamment par Craig et McCann (1978), Ferber (1966) et Francis et Bush (1995). Colsher et Wallace (1989) indiquent dans une étude portant sur les aînés la non-réponse à une question augmente avec l'âge. Ils suggèrent que les aînés sont peut-être peu familiers avec les méthodes de recherche utilisées et qu'ils hésitent peut-être à partager leurs expériences personnelles avec des étrangers. Toutefois, Ying (1989) observe que la non-réponse à une question est plus élevée chez les répondants déprimés et en mauvaise santé.

Les caractéristiques de l'intervieweur. Il n'est nulle part question, dans la documentation scientifique, des effets systématiques de caractéristiques particulières de l'intervieweur sur la non-réponse à une question. Par exemple, Berk et Bernstein (1988) observent un effet négatif de l'âge de l'intervieweur sur la non-réponse à une question tandis que Collins (1980) n'observe pas d'effet net de l'âge de l'intervieweur. Hox, De Leeuw et Kreft (1991) n'observent eux non plus aucun effet des caractéristiques de l'intervieweur sur la non-réponse à une question, même s'ils relèvent par ailleurs des différences de ce point de vue d'un intervieweur à l'autre. Toutefois, personne ne doute de

l'incidence que peuvent avoir les intervieweurs sur la non-réponse à une question (ils peuvent demander des éclaircissements, négliger de noter les résultats, etc.). En ce qui concerne les intervieweurs, on devrait probablement attacher plus d'importance à leur comportement lors de l'interview qu'à leurs caractéristiques intrinsèques. Dans la même veine, le style personnel de l'intervieweur (Rogers, 1976) et le fait qu'un intervieweur mentionne au répondant qu'il est difficile de répondre à certaines des questions (Sudman et coll., 1977) contribuent à accroître le taux de non-réponse à une question.

L'énoncé de la question. Certains indices portent à conclure à l'existence d'un rapport entre la difficulté de la question et la non-réponse à cette question. Dans une interview téléphonique, Leigh et Martin (1987) ont observé que plus une question était difficile, plus élevé était la non-réponse à la question. Dans une étude portant sur le bien-être, Ying (1989) a observé que les questions à formulation positive présentent des pourcentages de non-réponse à une question plus élevés. Dans le cadre d'une méta-analyse, Viswesvaran, Barrick et Onnes (1993) ont observé des effets cohérents de la difficulté d'une question sur le pourcentage de non-réponse à cette question. Par contre, Molenaar (1988) indique que malgré le lien apparent qui existe entre la difficulté d'une question et la non-réponse à cette question, rien ne permet d'affirmer que cette relation puisse être due à la difficulté linguistique des questions en cause (mesurée, par exemple, à l'aide d'une formule d'évaluation de la facilité de lecture).

Dans la présente étude, nous formulons un certain nombre de questions. Nous limitons notre analyse au temps, à certaines caractéristiques des répondants (sexe, âge, sensations de douleur, difficultés physiques éprouvées dans la vie de tous les jours), à la note du CES-D, à un certain nombre de variables liées au CES-D (par exemple, niveau de scolarité et environnement urbain du lieu de résidence), à certaines caractéristiques de l'intervieweur (âge et niveau de scolarité) et à l'énoncé des questions (à formulation positive ou négative).

La question examinée peut se résumer comme suit : quelle relations peut-on observer entre l'âge et la non-réponse à une question en fonction du temps dans l'échelle CES-D, dans des conditions connues de collecte des données et un certain nombre de caractéristiques pertinentes des répondants et des intervieweurs et l'énoncé de la question, en utilisant un modèle d'analyse multiniveau?

3. DONNÉES ET TRAVAIL SUR LE TERRAIN

Les données utilisées pour illustrer notre modèle sont tirées de l'étude LASA (*Longitudinal Aging Study Amsterdam*). La LASA est une étude longitudinale multidisciplinaire qui s'intéresse aux changements survenus dans l'autonomie et la qualité de vie des aînés, ainsi qu'aux prédicteurs et aux conséquences de ces changements (Deeg, Knipscheer et Van Tilburg, 1993). La cohorte de la LASA est un échantillon aléatoire stratifié pour l'âge, le sexe et la mortalité attendue sur une période de cinq ans. La base de sondage a été constituée à partir des registres de onze municipalités des régions de l'ouest, de l'est et du sud des Pays-Bas. Dans chaque strate, l'échantillon est représentatif de la population des personnes âgées (55 à 85 ans) des Pays-Bas. De septembre 1992 à septembre 1993, les participants de cette étude ont été interviewés dans le cadre de la première vague de la LASA (N = 3 107).

Les intervieweurs ont participé à un cours de formation visant à normaliser leur comportement (cinq séances de six heures chacune). On a eu recours à des exemples d'interviews filmées sur vidéo et à des jeux de rôle pour permettre aux intervieweurs de perfectionner leurs techniques. Des séances de formation supplémentaires ont été organisées pendant le travail sur le terrain. Les intervieweurs ont rempli un questionnaire afin d'évaluer leurs propres caractéristiques. À l'intérieur de chaque région, les répondants ont été répartis au hasard entre les intervieweurs. Toutes les interviews ont été enregistrées sur bande sonore afin de permettre un contrôle de la qualité des données. Elles ont toutes été conduites au domicile des répondants.

Dans l'interview de base, l'état dépressif léger a été mesuré à l'aide de l'échelle CES-D. Il s'agit d'un instrument d'auto-évaluation de 20 questions élaborées pour mesurer les symptômes de la dépression

au sein des communautés. Cette échelle, largement utilisée pour des échantillons de personnes âgées, a montré de bonnes propriétés psychométriques pour ces échantillons (Hertzog 1990; Radloff 1986). La version néerlandaise de cette échelle présentait des propriétés psychométriques semblables pour trois échantillons de personnes âgées des Pays-Bas (Beekman *et coll.*, 1994). L'absence d'un recoupement avec les symptômes de maladies physiques est cruciale pour l'étude des personnes âgées; on a démontré que ce recoupement était minimal dans un certain nombre d'études (Berkman *et coll.*, 1986). L'échelle CES-D génère une note totale qui peut varier de 0 à 60. Afin d'isoler les répondants pertinents au point de vue clinique, nous avons utilisé la note de césure de >15.

Après les interviews, 662 répondants ont été incorporés dans une étude de suivi de trois ans (six vagues séparées par des intervalles d'environ 150 jours). La base de données était constituée de l'ensemble des répondants ayant obtenu une note supérieure à 15 à l'échelle CES-D, et d'un échantillon aléatoire de répondants ayant obtenu une note entre 0 et 15. Le suivi a été réalisé à l'aide d'un questionnaire expédié par la poste et à remplir soi-même et qui permettait de déterminer la note à l'échelle CES-D. Avant le début de l'étude de suivi, sept des répondants ont été retirés de l'étude (pour cause de maladie, de décès ou d'autres raisons), laissant ainsi 655 répondants éligibles. Dans le présent article, nous présentons les données obtenues pendant l'étude de base et l'étude de suivi.

4. RÉSULTATS

Non-réponse à une question. Dans une enquête, la non-réponse est la proportion des répondants d'un échantillon visé par l'enquêteur qui n'ont pas fourni les données requises. La non-réponse totale d'une étude donnée peut être décomposée en une portion constituée des gens qui ont refusé de coopérer à l'étude («non-réponse relative à l'unité») et une autre constituée de personnes qui ont coopéré, mais qui n'ont pas donné une réponse utile à une question particulière («non-réponse à une question»). Si la non-réponse est aléatoire, elle sera sans effet sur la générabilité des résultats dans la population. Toutefois, dans le cas contraire, il conviendra de détecter les sources du biais. L'analyse de la non-réponse vise donc à détecter les sources et les facteurs qui contribuent au biais d'une façon significative. Si nous connaissons ces facteurs, nous pouvons tenter de corriger les résultats, par exemple, à l'aide de techniques d'imputation (telles que décrites par Little et Rubin, 1987). Toutefois, il paraîtrait plus sage d'éviter complètement le recours à de telles techniques en déterminant à l'avance les causes de la non-réponse à une question.

L'analyse de la non-réponse à une question doit être précédée d'une analyse de la non-réponse relative à l'unité, en particulier dans le cas d'une étude longitudinale. Dans le cas qui nous concerne, les taux de non-réponse relatives à l'unité s'établissaient à 20,5 %, 17,8 %, 31,1 %, 34,4 %, 40,7 % et 31,9 % pour les cycles d'enquête 2 à 7 respectivement. On a observé une baisse constante généralisée du taux de non-réponse relative à l'unité, sauf dans le cas du sixième cycle. Une analyse exploratoire de notre étude a démontré que la non-réponse relative à l'unité augmentait avec le temps, en même temps que l'âge des répondants et l'aggravation de leur état dépressif, et qu'elle diminuait en fonction du niveau de scolarité. Les hommes présentent des taux de non-réponse inférieurs à ceux des femmes. Le tableau 1 montre les taux de non-réponse relative à l'unité observés pour chaque cycle.

Non-réponse à une question. Nous présentons au tableau 1 les pourcentages de la non-réponse à une question pour chaque question. Les valeurs globales des taux de non-réponse à une question sont plutôt faibles (0,4; 3,0; 2,4; 1,4; 1,0; 1,0 et 0,6, pour les cycles 1 à 7 respectivement). Les pourcentages les plus faibles s'observent au cycle 1, lorsque les interviews ont été réalisées. En ce qui concerne les résultats du questionnaire transmis par la poste, on observe une baisse graduelle de la non-réponse à une question avec le temps.

Tableau 1. Pourcentage de la non-réponse à une question par question et par cycle

Cycle	1	2	3	4	5	6	7	
Nombre de répondants	655	521	540	517	495	388	446	moyenne
Question								
1	0,3	2,7	2,0	1,4	0,8	0,8	0,0	1,14
2	0,0	2,3	2,0	1,2	0,8	1,0	0,0	1,04
3	0,0	1,9	2,2	1,4	0,6	1,0	0,0	1,01
4	1,2	3,5	2,4	1,9	1,4	0,8	0,4	1,60
5	1,2	1,9	2,4	1,7	1,2	1,3	0,2	1,41
6	0,3	2,5	2,4	1,4	0,6	1,3	0,2	1,24
7	0,2	2,5	2,6	1,7	0,6	1,3	0,2	1,30
8	2,5	4,2	3,2	2,3	2,0	1,0	0,1	2,19
9	0,2	3,5	3,2	2,1	1,1	1,3	0,7	1,73
10	0,2	3,5	2,6	2,1	1,0	1,0	0,3	1,53
11	0,2	2,5	2,0	1,4	0,8	0,8	1,0	1,24
12	1,0	4,0	2,8	2,1	1,6	1,0	1,1	1,94
13	0,3	4,2	3,0	1,7	1,1	1,3	1,6	1,89
14	0,0	1,7	1,3	1,2	1,2	1,3	1,1	1,11
15	0,0	3,5	3,2	1,2	1,2	1,3	1,1	1,64
16	0,5	3,8	2,2	1,0	0,8	0,5	0,9	1,39
17	0,2	3,1	2,4	1,0	1,0	0,8	1,1	1,37
18	0,3	2,3	2,6	1,0	1,4	0,8	1,3	1,39
19	0,2	4,4	2,4	0,4	1,2	1,0	0,8	1,37
20	0,0	3,3	1,3	0,8	0,6	1,0	0,8	1,11

Méthodologie. La plupart des variables de l'étude avaient une distribution asymétrique. Pour corriger ce problème, nous avons utilisé les valeurs de l'écart réduit pour variables de type intervalle aux fins de l'analyse. Ces valeurs n'étaient pas centrées par rapport aux niveaux.

Comme la variable dépendante est dichotomique et que les variables indépendantes sont de type intervalle, nous avons utilisé des modèles de régression logistique pour une réponse binaire (Goldstein, 1991). Dans un tel modèle, le logit doit être prédit. Dans la présente étude, ce logit correspond au logarithme naturel du rapport de la probabilité de la non-réponse à une question sur la probabilité d'une réponse à cette question. Les modèles ont été analysés à l'aide du programme MLN (Woodhouse et coll., 1995).

Les questions de recherche ont été traduites en un modèle de coefficient aléatoire d'analyse multiniveau qui, sous sa forme la plus simple, consiste en deux équations. La première équation (niveau 1 : cycles d'enquête) correspond à la régression interne au cycle où le logit du taux de non-réponse par question (INR), considéré comme une variable dépendante, est expliqué par le temps. En ce qui concerne un cycle i pour un répondant j , on obtient l'équation suivante :

$$\ln(\text{INR}_{ij}/(1-\text{INR}_{ij})) = \beta_{0j} + \beta_{1j} t_{ij} + e_{ij} \quad (1)$$

où i désigne les cycles d'enquête ($i = 1, \dots, 7$), j désigne le répondant ($j = 1, \dots, J$), $\ln(\text{INR}_{ij}/(1-\text{INR}_{ij}))$ désigne la variable dépendante, β_{0j} représente la moyenne pour le cycle i , et β_{1j} représente le coefficient de régression de la variable dépendante $\ln(\text{INR}_{ij}/(1-\text{INR}_{ij}))$ pour t_{ij} . Les variables dépendantes négligées par erreur et les erreurs de mesure (notamment les effets des caractéristiques du mode, du répondant et de l'intervieweur, et l'énoncé des questions) sont exprimées dans le terme d'écart (e_{ij}). $e_{ij} \sim N(0, \sigma_e^2)$. Les hypothèses habituelles de l'analyse de régression s'appliquent ici.

Au niveau du répondant, l'équation se présente comme suit :

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} Z_{1j} + \dots + \gamma_{kN} Z_{Nj} + u_{kj} \quad (2)$$

où Z_{nj} désigne les variables indépendantes ($n=1, \dots, N$) du répondant j , γ_{k0} ($k=0, 1$) désigne l'ordonnée à l'origine de la droite de régression, γ_{kn} désigne le coefficient de régression de la variable dépendante β_{kj} sur la variable indépendante du répondant Z_{nj} , et u_{kj} ($u_{kj} \sim 0, \sigma_u^2$) désigne le terme d'écart correspondant à β_{kj} . $E(e_{ij}u_{kj})=0$. La substitution de l'équation (2) dans l'équation (1) donne une

nouvelle équation où les termes qui contiennent γ_{kn} désignent la portion fixe du modèle et ceux qui contiennent e_{ij} ou u_{kj} désignent la portion aléatoire. Cette méthode présente certaines similitudes avec celle de Singh et coll., (1996), quoique ces derniers n'aient pas contrôlé le niveau intermédiaire des répondants.

Dans l'analyse, les intervieweurs et les questions sont ajoutés à titre de troisième niveau; les équations correspondantes se comportent de façon analogue. Nous présentons dans le tableau 2 l'ordonnancement hiérarchique des données servant de fondement à l'analyse multiniveau. Puisque les cycles d'enquête sont subordonnés aux répondants, il est possible d'omettre les cas présentant des valeurs manquantes de la variable dépendante dus à la non-réponse relative à l'unité. Tous les cycles possibles sont présentés dans le cas du répondant 1, mais non dans le cas du répondant 655.

Tableau 2. Structure hiérarchique des données (exemple fictif). Non-réponse à une question = 1; réponse à une question = 0. Les cas de valeurs manquantes de la variable dépendante dus à la non-réponse relative à l'unité sont exclus du fichier de données.

cycle numéro	répondant numéro	intervieweur numéro	question numéro	réponse à la question k
1	1	1	1	0
2	1	1	1	1
3	1	1	1	0
4	1	1	1	0
5	1	1	1	0
6	1	1	1	0
7	1	1	1	0
...
1	655	24	20	0
3	655	24	20	1
4	655	24	20	0
5	655	24	20	0
7	655	24	20	0

Effets du temps et effets du mode. Outre l'analyse de la réduction de la non-réponse totale dans le temps, nous avons procédé à une analyse par question. Dans cette analyse, nous testons l'hypothèse selon laquelle la non-réponse à une question est plus élevée dans le cas des interviews directes que dans celui des questionnaires expédiés par la poste (l'effet du mode). Les résultats de cette analyse sont présentés au tableau 3

Tableau 3. Logit de la moyenne de la non-réponse à une question et régressions du TEMPS et du MODE. Les valeurs des écarts-types sont indiquées entre parenthèses.

Question	MOYENNE	TEMPS	MODE
1	-3,91 (0,44)	-0,45 (0,12)	1,32 (0,39)
2	-7,76 (10,98)	-0,35 (0,11)	4,77 (10,98)
3	-7,70 (9,00)	-0,37 (0,12)	4,75 (9,01)
4	-3,32 (0,29)	-0,37 (0,10)	0,75 (0,52)
5	-4,64 (0,56)	-0,25 (0,10)	1,49 (0,23)
6	-4,09 (0,43)	-0,36 (0,11)	1,25 (0,38)
7	-4,12 (0,56)	-0,35 (0,10)	1,62 (0,52)
8	-2,95 (0,23)	-0,29 (0,08)	0,41 (0,18)
9	-4,33 (0,54)	-0,32 (0,09)	1,73 (0,52)
10	-4,49 (0,54)	-0,26 (0,09)	1,64 (0,51)
11	-4,67 (0,56)	-0,25 (0,10)	1,47 (0,52)
12	-3,46 (0,29)	-0,29 (0,08)	0,83 (0,24)
13	-4,06 (0,41)	-0,27 (0,08)	1,36 (0,37)
14	-7,63 (16,24)	-0,08 (0,98)	4,62 (16,24)
15	-7,22 (7,91)	-0,29 (0,09)	4,44 (7,91)
16	-3,68 (0,37)	-0,42 (0,10)	1,18 (0,32)
17	-4,52 (0,55)	-0,29 (0,10)	1,57 (0,52)
18	-4,42 (0,42)	-0,19 (0,09)	1,08 (0,38)
19	-4,23 (0,55)	-0,40 (0,09)	1,76 (0,52)
20	-7,61 (6,44)	-0,32 (0,11)	4,55 (6,44)

Le tableau montre que tous les effets du temps sont négatifs; ainsi, la non-réponse à une question tend généralement à diminuer pour l'ensemble des questions. Ces résultats viennent appuyer l'hypothèse de l'effet du mode : les intervieweurs obtiennent des taux de non-réponse aux questions inférieurs à ceux obtenus à l'aide des questionnaires expédiés par la poste.

Ces résultats peuvent se prêter à l'interprétation suivante. Nous utilisons la question 1 en guise d'exemple. La valeur du logit moyen est -3,91, ce qui signifie qu'en moyenne, 4,9 % [=100*exp^{-0,391}/(1+exp^{-0,391})] des répondants tendent à refuser de donner une réponse à la question. Dans le cas des interviews, ce pourcentage est de 2,7 % [=100*exp^{-0,391+1,32}/(1+exp^{-0,391+1,32})], et au point médian fictif de la série de sondages postaux, il est de 5,9 % [=100*exp^{-3,91-1,32}/(1+exp^{-3,91-1,32})]. Au troisième cycle, qui représente le cycle médian dans le temps, ce pourcentage est de 5,5 % [=100*exp^{-3,91-0,45+1,32}/(1-exp^{-3,91-0,45+1,32})].

Effets du répondant et de l'intervieweur. L'étape suivante de l'analyse consiste à évaluer les différences qui existent dans la fonction temps et de chercher à expliquer ces différences au moyen des variables qui correspondent au répondant. Le modèle de l'analyse peut être caractérisé comme suit. Les cycles, les répondants et les intervieweurs ont été incorporés dans l'analyse sous forme de niveaux distincts. Les variables du répondant mentionnées ont également été incorporées dans le modèle. Toutefois, nous n'avons observé en aucun cas une différence significative due à l'intervieweur. Si nous limitons notre analyse au premier cycle, celui des interviews, les différences sont très légères; la valeur de η^2 est en général de 0,03.

Nous présentons au tableau 4 les résultats de cette analyse pour chaque question.

Tableau 4. Coefficients fixes significatifs des effets de la MOYENNE, du TEMPS, du SEXE et de l'ÂGE sur le logit de la non-réponse à une question (écarts-types non indiqués).

Question	MOYENNE	TEMPS	SEXE	ÂGE
1	-5,22	-0,47	1,48	0,81
2	-5,84	-0,36	1,34	0,72
3	-5,60	-0,38	1,34	0,77
4	-4,27	-0,37	0,69	0,62
5	-4,89	-0,26	0,77	0,58
6	-4,27	-0,37	0,46	0,78
7	-4,91	-0,36	0,92	0,67
8	-3,96	-0,29	0,57	0,58
9	-4,72	-0,33	0,89	0,81
10	-5,31	-0,28	1,15	0,79
11	-5,30	-0,36	0,97	0,77
12	-5,24	-0,39	1,24	0,73
13	-5,63	-0,27	1,32	0,89
14	-5,64	-0,00	0,83	0,58
15	-5,13	-0,29	1,03	0,82
16	-5,60	-0,42	1,37	0,80
17	-5,53	-0,30	1,13	0,84
18	-5,16	-0,18	0,86	0,65
19	-4,61	-0,41	0,76	-1,01
20	-5,01	-0,32	0,00	0,70

Comme il existe une intercorrélation, quoique légère, entre le temps et le mode et que les résultats concernant l'hypothèse de l'effet du mode sont tellement évidents, nous avons décidé d'omettre les données obtenues par voie d'interview. Le nombre de cycles est ainsi réduit à six.

À partir de cette base de données modifiée, nous avons procédé à une analyse multiniveau sur un échantillon de toutes les questions. Les résultats sont présentés au tableau 5. Compte tenu des valeurs significatives de la portion aléatoire du modèle 1, on peut conclure que les trajectoires de la non-réponse à une question diffèrent d'un répondant à l'autre. Comme prévu, le coefficient lié au temps est négatif. Dans le modèle 2, toutes les variables correspondant au répondant mentionnées ci-dessus ont été incorporées. Dans la portion fixe, les effets du sexe et de l'âge ont disparu et ceux liés aux difficultés dans la vie de tous les jours (DIFF), à l'état dépressif (CES-D) et à l'expérience de la douleur (PAIN) ont laissé voir des effets significatifs. Le niveau de scolarité, la pratique religieuse, l'environnement urbain et la santé subjective absolue et relative n'ont laissé voir aucun effet. Les trajectoires spécifiques au répondant demeurent.

Effets de l'énoncé de la question. Nous nous sommes en outre demandé s'il existait ou non des différences entre les questions, selon leur type de formulation (positive ou négative). Pour déterminer

l'existence d'un effet possible, nous avons ajouté les questions pour en faire un troisième niveau et utiliser la formulation négative ou positive comme une variable. Le modèle 3 du tableau 5 montre qu'en plus de l'effet des variables du répondant, les questions à formulation négative présentent une probabilité plus faible de non-réponse que celles à formulation positive.

Tableau 5. Analyse multiniveau de trois modèles sur toutes les questions pour six cycles d'enquête. La variable dépendante est le logit lié à la réponse à la question. Dans le modèle 1, il n'y a pas de contrôle des variables du répondant; dans le modèle 2, toutes les variables du répondant sont introduites. Dans le modèle 3, la question est ajoutée pour donner un troisième niveau. Les valeurs des écarts-types sont indiquées entre parenthèses; les coefficients non significatifs sont omis.

	Modèle 1	Modèle 2	Modèle 3
<i>Portion fixe</i>			
CONS	-0,98 (0,17)	-2,01 (0,13)	-1,40 (0,11)
TEMPS	-0,94 (0,10)	-1,31 (0,08)	-0,70 (0,04)
DIFF		5,19 (0,16)	-0,16 (0,06)
CES-D		0,37 (0,06)	0,54 (0,05)
PAIN		3,92 (0,13)	-0,43 (0,05)
NEG/POS			-1,30 (0,15)
<i>Portion aléatoire</i>			
Niveau du cycle			
CONS/CONS			0,00 (0,00)
Niveau du répondant			
CONS/CONS	63,86 (2,82)	15,34 (1,59)	6,24 (0,86)
CONS/TEMPS	-42,73 (1,70)	-14,62 (1,08)	-2,44 (0,38)
TEMPS/TEMPS	27,62 (1,04)	11,68 (0,74)	0,16 (0,16)
Niveau du temps			
CONS/CONS	1 (0)	1 (0)	1 (0)

6. CONCLUSION

Dans le présent article, nous nous sommes surtout penchés sur la façon d'analyser la trajectoire de la non-réponse à une question, en tenant compte de cinq sources possibles de biais. Afin d'éclaircir quelque peu cette question, nous avons examiné les aspects suivants : a) Quel rapport existe-t-il entre la non-réponse à une question et le temps? Dans notre étude, la non-réponse à une question a montré une tendance à la baisse. On peut supposer que les répondants capables et motivés ont tendance à poursuivre leur collaboration. b) Quelles différences existe-t-il entre les non-réponses à une question obtenues avec différents modes de collecte des données? Conformément à nos prévisions, nous avons observé que les entrevues directes présentent des taux de non-réponse inférieurs aux questionnaires postés à remplir soi-même. c) La non-réponse à une question dépend-elle des caractéristiques du répondant? Nous avons prévu que le sexe, l'âge, le niveau de scolarité et la santé des répondants auraient une incidence sur les taux de non-réponse à une question. Conformément aux conclusions des autres recherches portant sur cette question, nous avons observé que le taux de non-réponse à une question augmente en fonction de l'âge et qu'il est plus élevé dans le cas des femmes. Notre analyse a en outre révélé que le sexe et l'âge semblent effectivement avoir un effet sur le taux de non-réponse, mais que la prise en compte de la santé physique et des variables liées à l'état dépressif montrent que la non-réponse à une question peut être expliquée par ces variables; nous avons en outre déterminé que le niveau de scolarité, l'environnement urbain et la santé subjective n'ont aucun effet. Ces résultats signifient que l'hypothèse de la non-familiarité avec l'enquête proposée par Colsher et Wallace (1989) n'est pas vérifiée. Il semble que l'état physique et l'état dépressif soient des variables qui expliquent mieux la non-réponse à une question. Par ailleurs, nous avons observé des différences considérables des trajectoires de la non-réponse en fonction du temps. d) La non-réponse à une question dépend-elle de l'intervieweur? Nous n'avons pas observé de telles différences. e) Existe-t-il des différences dues à la formulation des questions? Nos résultats montrent que les taux de non-réponse sont moindres dans le cas des questions à formulation positive que dans le cas des questions à formulation négative.

Nous avons proposé une méthode d'évaluation des effets du temps et des caractéristiques du répondant et de l'intervieweur sur le taux de non-réponse à une question dans une enquête longitudinale. Nous avons montré que la subordination des cycles d'enquête aux répondants et des répondants aux intervieweurs nécessitent l'utilisation d'un modèle multiniveau. Comme la non-réponse à une

question est une variable dichotomique, il convient d'utiliser une analyse de régression logistique à niveaux multiples. Par conséquent, nous présentons une description générale d'un modèle adéquat pour l'étude de ces effets puisqu'ils tirent leur origine de sources à divers niveaux. Nous avons finalement montré l'existence d'effets dus au temps sur la non-réponse à une question, dans une étude longitudinale. Notre analyse a démontré l'existence d'effets dus au mode, dans la mesure où une interview produit des taux de non-réponse inférieurs à ceux obtenus à l'aide d'un questionnaire expédié par la poste. Les effets dus à l'âge des répondants n'ont pu être observés, mais il semble que les difficultés physiques éprouvées dans la vie de tous les jours, l'état dépressif et l'expérience de la douleur soient les déterminants principaux.

Même si les pourcentages de la non-réponse à une question sont en général faibles, il peut être utile de les expliquer même si le processus est laborieux. L'attention portée aux détails de la collecte des données est la première précaution à prendre pour accroître la qualité des données des enquêtes.

7. BIBLIOGRAPHIE

Beekman, A.T.F., Van Limbeek, J., Deeg, D.J.H., Wouters, L. et Van Tilburg, W. (1994). Een screeninginstrument voor depressie bij ouderen in de algemene bevolking: de bruikbaarheid van de Center For Epidemiologic Studies Depression Scale (Ces-D). (Un instrument de dépistage de l'état dépressif parmi les personnes âgées de la population générale : utilité de l'échelle CES-D). *Tijdschrift voor Gerontologie en Geriatrie*.

Berk, M.L. et Bernstein, A.B. (1988). Interviewer characteristics and performance on a complex health survey. *Social Science Research*, 17, 239-251.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. et Sudman, S. (Eds.). (1991). *Measurement Errors in Surveys*. New York: John Wiley and Sons.

Collins, M. (1980). Interviewer variability: A review of the problem. *Journal of the Market Research Society*, 77-95.

Colsher, P.L. et Wallace, R.B. (1989). Data quality and age: health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology*, 44, 45-52.

Craig, C.S. et McCann, J.M. (1978). Item nonresponse in mail surveys: extent and correlates. *Journal of Marketing Research*, 15, 285-289.

Deeg, D.J.H., Knipscheer, C.P.M. et Van Tilburg, W. (1993). *Autonomy and Well-being in the Aging Population: Concepts and Design of the Longitudinal Aging Study Amsterdam*. Bunnik: NIG.

De Leeuw, E. (1992). *Data Quality in Mail, Telephone, and Face to Face Surveys*. Amsterdam, T-Publications.

Ferber, R. (1966). Item nonresponse in a consumer survey. *Public Opinion Quarterly*, 30, 399-415.

Francis, J.D. et Bush, L. (1975). What we know about "I do'nt knows". *Public Opinion Quarterly*, 34, 207-218.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78, 45-51.

Groves, R.M. (1989). *Survey Error and Survey Cost*. New York etc: John Wiley and Sons.

Groves, R.M. et Fultz, R.L. (1979). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research*, 14, 31-52.

Hertzog, C, Van Alstine, J., Usala, P.D., Hultsch, D.F. et Dixon, R. (1990). Measurement proportions of the CES-D scale in older populations. *Psychological Assessment*, 2, 64-72.

Hox, J.J. De Leeuw, E.D. et Kreft, I.G.G (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In: P.P. Biemer, R.M Groves, L. E. Lyburg, N.A. Mathiowetz, et S. Sudman (eds). *Measurement Errors in Surveys*. New York: Wiley, 439-462.

Leigh, J.H. et Martin, Cl.R. (1987). "Don't know" item nonresponse in a telephone survey: Effects of question form and respondent characteristics. *Journal of Marketing Research*. 24, 418-424.

Little, D.B. et Rubin, R.J.A. (1987). *Statistical Analysis with Missing Data*. New York, etc. Wiley.

Madow, W.G., Nisselson, H. et Olkin, I. (1983). *Incomplete Data in Sample Surveys, Vol 1: Report and Case Studies*. New York: Academic Press.

Molenaar, N.J. (1988). *Formulerings-effecten in Survey-interviews: een Non-experimenteel Onderzoek*. (Question Wording in Survey-Interviews). Amsterdam: Vrije Universiteit.

Radloff, L.S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 3, 385-401.

Rogers, Th.F. (1976). Interviews by telephone and in person: Quality of responses and field performance. *Public Opinion Quarterly*, 42, 407-410.

Singh, A.C. et Merkouris, P. (1996). A diversity component analysis of interviewer effects in categorical survey data. In: SMPQ (eds.) *Proceedings of the International Conference on Survey Measurement and Data Quality*. American Statistical Association, Alberta, Vi. 332-337.

Sudman, S., Bradburn, N.M., Blair, E. et Stocking, C. (1977). Model expectation: The effect of interviewer's prior expectations on responses. *Sociological Methods and Research*, 6, 171-182.

Viswesvaran, Ch., Barrick, M.R. et Ones, D.S. (1993). How definitive are conclusions based on survey data: Estimating robustness to nonresponse. *Personel Psychology*, 46, 551-567.

Woodhouse, G. (eds). (1995). *A Guide to Mln for New Users*. London, Multilevels Projects, Institute of Education, University of London.

Ying, Y.W. (1989). Nonresponse to the Center for Epidemiological Studies Depression scale in Chinese Americans. *International Journal of Social Psychiatry*, 35, 156-163.

ANNEXE 1 : QUESTIONS DE L'ÉCHELLE CES-D D'ÉVALUATION DE L'ÉTAT DÉPRESSIF

Prière d'indiquer combien de fois au cours de la dernière semaine vous vous êtes senti comme cela (0: rarement ou jamais, 1: quelquefois, 2: de temps en temps 3: souvent ou toujours)

1. Au cours de la dernière semaine, je me suis inquiété à propos de choses qui ne m'inquiètent pas habituellement.
2. Au cours de la dernière semaine, j'ai manqué d'appétit.
3. Au cours de la dernière semaine, j'ai eu l'impression de ne pouvoir me sortir de ma déprime, même avec l'aide de ma famille et de mes amis.
4. Au cours de la dernière semaine, j'ai eu l'impression d'être aussi bon que les autres.
5. Au cours de la dernière semaine, j'ai eu du mal à me concentrer sur ce que je faisais.
6. Au cours de la dernière semaine, je me suis senti déprimé.
7. Au cours de la dernière semaine, j'ai eu l'impression que tout ce que je faisais demandait un effort.
8. Au cours de la dernière semaine, je me suis senti confiant en l'avenir.
9. Au cours de la dernière semaine, j'ai pensé que ma vie était un échec.
10. Au cours de la dernière semaine, je me suis senti craintif.
11. Au cours de la dernière semaine, j'ai eu un sommeil agité.
12. Au cours de la dernière semaine, je me suis senti heureux.
13. Au cours de la dernière semaine, j'ai parlé moins que d'habitude.
14. Au cours de la dernière semaine, je me suis senti seul.
15. Au cours de la dernière semaine, les gens m'ont paru hostiles.
16. Au cours de la dernière semaine, j'ai apprécié la vie.
17. Au cours de la dernière semaine, j'ai eu des crises de larmes.
18. Au cours de la dernière semaine, j'ai éprouvé de la tristesse.
19. Au cours de la dernière semaine, j'ai eu l'impression que les gens ne m'aimaient pas.
20. Au cours de la dernière semaine, j'ai manqué d'énergie.

ÉVALUATION DE L'ERREUR NON DUE À L'ÉCHANTILLONNAGE DANS LES DONNÉES D'ENQUÊTE À L'AIDE DE MODÈLES AVEC ORDONNÉE À L'ORIGINE ALÉATOIRE

Dale Atkinson¹

RÉSUMÉ

La variation de la déclaration faite par un répondant d'une période d'enquête à l'autre a de nombreuses causes. Ainsi, la variation déclarée des stocks reflète l'effet conjugué du changement réel et de l'erreur non due à l'échantillonnage. La présente étude vise à 1) déterminer les composantes systématiques de la variabilité de la déclaration d'une enquête à l'autre qui pourraient être le signe d'une erreur non due à l'échantillonnage, 2) quantifier l'erreur non due à l'échantillonnage éventuellement sous-jacente et 3) repérer les domaines où des mesures correctives pourraient être prises en vue de limiter le problème.

On recourt à des modèles avec ordonnée à l'origine aléatoire pour quantifier la fraction de la variabilité d'une enquête à la suivante expliquée par les effets différentiels de la tâche de l'enquêteur et du changement de répondant. Sous certaines hypothèses, la modélisation fournit une estimation de diverses mesures normalisées de la qualité de l'enquête, comme la fiabilité. Quand on intègre la tâche de l'enquêteur au modèle, la méthode permet de repérer les tâches qui contribuent de façon disproportionnée à la variabilité. Elle offre donc un moyen de dépister les problèmes de dénombrement éventuels qui pourraient justifier de la formation conceptuelle supplémentaire. Ainsi, de nombreux renseignements sur le processus de collecte des données peuvent être tirés directement des données d'enquête, sans qu'il soit nécessaire de recueillir de nouvelles données.

Le présent article traite de modèles avec ordonnée à l'origine aléatoire, des résultats de l'application de cette méthode à des données recueillies récemment par le *National Agricultural Statistics Service (NASS)* du *U.S. Department of Agriculture*, et de la possibilité d'intégrer entièrement cet instrument de contrôle de la qualité au processus de gestion d'enquête du NASS.

MOTS CLÉS: Enquête par sondage; erreur non due à l'échantillonnage; variabilité de la déclaration; modèles avec ordonnée à l'origine aléatoire.

1. INTRODUCTION

1.1 Description du problème

Malgré les efforts déployés pour les éviter, des erreurs non dues à l'échantillonnage entachent les données d'enquête. Ces erreurs, dont il existe de nombreuses formes, ont une incidence sur les données à diverses étapes du processus d'enquête. Elles résultent d'activités accomplies avant, pendant et après l'interview. Le présent article traite des erreurs qui surviennent durant l'interview et vise surtout à déterminer s'il est possible de déceler la composante systématique (et, avec un peu de chance, corrigible) de la variabilité de la déclaration.

Les erreurs commises au moment de l'interview peuvent être imputables à l'intervieweur, au répondant ou à une interaction entre les deux. Le changement de répondant est une source éventuelle d'écart entre les déclarations trimestrielles. L'effet différentiel d'intervieweur sur la réponse en est une autre. On examine ici

¹Dale Atkinson, Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, 3251 Old Lee Hwy, Room 305, Fairfax, Va., U.S.A., 22030.

ces deux possibilités, afin de déterminer si l'une ou l'autre a une incidence significative sur les données recueillies.

2. CONTEXTE

Le *National Agricultural Statistics Service (NASS)* du *U.S. Department of Agriculture (USDA)* effectue un cycle permanent d'enquêtes trimestrielles par sondage pour estimer les stocks et la production de divers produits agricoles. Ces enquêtes trimestrielles ont lieu en juin, en septembre, en décembre et en mars. Le plan d'échantillonnage fait appel à des bases de sondage multiples, soit une combinaison de listes et de bases aréolaires. En juin, on dénombre complètement les segments aréolaires échantillonnés, dont la taille est de l'ordre d'un mille carré, afin d'enregistrer toutes les activités agricoles qui ont lieu dans leurs limites. Les arrangements concernant l'exploitation des terres (secteurs) dans les segments échantillonnés définissent les unités pour lesquelles on recueille les données d'enquête. Les données provenant de ces secteurs fournissent des estimations complètes pour les bases aréolaires, ainsi que la composante aréolaire des mesures fondées sur la base de sondage multiple. Ces mesures n'incluent des données aréolaires que pour les secteurs aréolaires dont la probabilité de sélection à partir des listes est nulle. Ces secteurs sont appelés secteurs non chevauchants (SNC).

Les SNC repérés en juin sont sous-échantillonnés pour tenir compte du caractère incomplet des listes de la base de sondage multiple utilisées pour recueillir les données des enquêtes trimestrielles de septembre, de décembre et de mars, lorsqu'on n'effectue pas de dénombrement aréolaire complet. La liste de bon nombre des bases de sondage utilisées par le NASS pour les divers produits étant assez complète, le nombre de secteurs non chevauchants (SNC) est relativement faible. Par exemple, en juin 1995, la variation du nombre de SNC selon l'état allait d'environ 10 à environ 300. Cependant, même si leur nombre est peu élevé, les SNC sont à l'origine d'une fraction importante de la variation des données obtenues à partir d'une base de sondage multiple. En conséquence, on veille à ce que leur taux d'échantillonnage soit élevé au moment des enquêtes trimestrielles subséquentes, de façon à obtenir pour la plupart des mesures répétées pour certaines questions.

Selon diverses analyses des données des enquêtes trimestrielles du NASS effectuées ces dernières années, la variabilité de réponse, de trimestre en trimestre, à des questions pour lesquelles la réponse devrait demeurer relativement stable durant l'année de référence est considérable pour certaines unités d'échantillonnage. Des enquêtes par réinterview, visant à déterminer la «valeur réelle», ont été effectuées dans le passé par le NASS pour résoudre des problèmes de déclaration similaires. Cependant, ces enquêtes imposent aux répondants un fardeau supplémentaire, situation problématique, compte tenu du fardeau déjà lourd lié aux opérations visant les SNC. En vue d'évaluer la qualité des données et de repérer les régions où des améliorations sont nécessaires, il serait donc préférable (si cette méthode donne de bons résultats) de recueillir autant de renseignements que possible sur la qualité de l'enquête auprès des personnes ordinairement chargées d'effectuer cette enquête. Cette approche est celle étudiée ici.

3. MÉTHODE

3.1 Renseignements généraux

Très souvent, quand on étudie l'erreur non due à l'échantillonnage, une source administrative ou une enquête par réinterview fournit les «valeurs réelles» qui permettent d'évaluer la qualité des données d'enquête. Dans le cas de notre étude, aucune de ces sources n'est disponible. Toutefois, comme le NASS déploie des ressources importantes en vue d'assurer la qualité des données collectées durant l'enquête sur l'agriculture de juin, on peut argumenter que ces données sont de meilleure qualité que celles recueillies lors des trimestres subséquents. Essentiellement, les données de juin possèdent nombre des attributs des données de réinterview fournissant la «valeur réelle». Les données de juin sont recueillies au cours d'interviews sur place (soi-disant le meilleur mode de collecte de données) par les enquêteurs du NASS les mieux formés et les plus chevronnés en matière d'agriculture. En revanche, la plupart des données des enquêtes trimestrielles subséquentes sont recueillies par téléphone, des bureaux locaux de la statistique du NASS. Les enquêteurs qui effectuent les

interviews par téléphone sont habituellement engagés sur place et, souvent, connaissent moins bien les enquêtes sur l'agriculture que les enquêteurs de terrain assignés en juin.

De toute évidence, aussi bien les données trimestrielles de juin que celles des enquêtes subséquentes contiennent des erreurs. Cependant, si les données de juin contiennent moins d'erreurs et représentent un substitut raisonnable des données réelles, il devient possible d'estimer les mesures de la qualité des données, comme la fiabilité et les indices d'incohérence, directement à partir des données trimestrielles. La méthode de régression avec ordonnée à l'origine aléatoire fournit, en outre, des estimations de la corrélation à l'intérieur des groupes et des effets de divers niveaux d'une variable de groupement (par exemple, tâche de l'enquêteur durant les enquêtes subséquentes ou changement éventuel de répondant entre deux sondages trimestriels consécutifs). Les pourcentages de la variabilité du modèle (quant à la prédiction de la réponse à une enquête trimestrielle subséquente d'après la réponse de juin) attribuable à des effets de groupe différentiels ont été calculés.

3.2 Le modèle

Biemer et Atkinson (1995 (1 et 2)) décrivent une méthode qui permet d'obtenir les mesures de la qualité des données et des effets de groupe pour des variables de groupement arbitraires à partir d'échantillons à deux degrés; l'échantillon de second degré étant sélectionné aux fins de la réinterview avec rapprochement des données pour obtenir les valeurs réelles. On essaie ici d'appliquer leur méthode au cas où les données de réinterview n'existeraient pas, mais où des mesures répétées indépendantes ont été effectuées sur les unités d'échantillonnage durant le cycle d'enquêtes trimestrielles. Dans le cas de ces enquêtes, on prévoit qu'une réponse (celle de juin) est, en général, meilleure que les autres et représente un «substitut de la valeur réelle» raisonnable. Le développement du modèle sous-jacent étant décrit en détail dans les références susmentionnées, on ne présentera pas ici un exposé aussi rigoureux. En général, on ajuste un modèle de la forme suivante :

$$y_i = \gamma_0 + \gamma \mu_i + z_{gi} \quad (3.2.1)$$

où μ_i est la valeur de juin de la variable observée, γ_0 et γ sont des constantes, et z_{gi} est un terme d'erreur aléatoire. Dans la mesure où μ_i est la valeur «réelle» de la variable observée, le paramètre γ_0 peut être interprété comme une constante, ou biais absolu, ajoutée à toutes les observations, tandis que γ représente un biais «proportionnel». À titre d'exemple, supposons que μ_i représente une mesure de la taille de l'entreprise agricole (par exemple, la superficie totale des terres de l'entreprise ou la superficie de terres labourables). La grandeur de l'erreur qui entache y_i est souvent proportionnelle à la taille, donc, modélisée comme il convient par le terme $\gamma \mu_i$. Le terme z_{gi} représente la somme de deux composantes aléatoires, d_g et δ_i , où d_g représente le «biais» ou un «effet de groupe» associé au groupe g , et δ_i représente une erreur indépendante de niveau unitaire. Nous supposons que $d_g \sim (0, \sigma_d^2)$ et $\delta_i \sim (0, \sigma_\delta^2 \mu_i^\lambda)$, où λ est une constante connue. Dans la présente étude, on a donné à λ la valeur de 0; cependant, il est possible d'estimer λ à partir des données (consulter, par exemple, Wright, 1983).

Considérons le modèle susmentionné et supposons en outre que la covariance conditionnelle des erreurs pour ($i \in G_g$) est donnée par

$$\begin{aligned} \text{Cov}(z_{gi}, z_{g'i'} | i) &= \sigma_d^2 + \sigma_\delta^2 \mu_i^\lambda \quad \text{pour } i = i' \\ &= \sigma_d^2 \quad \text{pour } i \neq i', g' = g \\ &= 0 \quad \text{pour } g' \neq g \end{aligned}$$

Représentons par $E(\bullet | i)$ l'espérance conditionnelle, étant donné l'unité i , de la distribution de l'erreur de mesure et par $\text{Var}(\bullet)$, la variance non conditionnelle de la distribution d'échantillonnage. Si on suppose que toutes les erreurs $G_g, g = 1, \dots, J$ sont de même taille (disons m) et que la correction pour une population finie

est négligeable, alors Biemer et Stokes (1991) montrent que $n\text{Var}(\bar{y}) = R^{-1}\text{Var}(\gamma\mu_i)[1 + (m-1)\rho_y]$, où R , appelé ratio de fiabilité, est donné par

$$R = \frac{\text{Var}E(y_i | i)}{\text{Var}(y_i)} = \frac{\gamma^2 \sigma_\mu^2}{\sigma_y^2}, \quad (3.2.2)$$

et où (quand on utilise une variable de groupement pour la tâche de l'enquêteur) ρ_y , appelé *coefficient de corrélation propre à l'enquêteur*, représente la corrélation entre des paires d'unités qui font partie de la tâche de l'enquêteur.

Le ratio de fiabilité, R , est le ratio de la variance de la valeur «réelle» de la variable observée, soit *viz.*, $\text{Var}(\gamma_0 + \gamma\mu_i)$, à la variance de l'observation y_i . Pour estimer R , il est habituellement nécessaire d'obtenir des mesures répétées, effectuées dans des conditions d'enquête identique et telles que les erreurs associées à chacune soient indépendantes (entre mesures) et aient une même distribution (consulter Biemer et Stokes, 1991). L'exécution d'une enquête par réinterview bien conçue est peut-être le meilleur moyen de satisfaire ces hypothèses, mais la méthode décrite ici permet d'estimer R directement à partir des données des enquêtes trimestrielles.

Étant donné le modèle (3.2.1), le coefficient de corrélation propre à l'enquêteur, ρ_y , est donné par

$$\rho_y = \frac{\sigma_d^2}{\sigma_y^2}. \quad (3.2.3)$$

Il est fort courant, lors des études de l'erreur de mesure, de décrire le degré de variation de la qualité de l'interview en fonction de l'enquêteur (consulter, par exemple, Groves, 1989).

Dans le cas de la présente étude, on estime ρ_y en prenant pour variable de groupement la tâche de l'enquêteur pour l'enquête subséquente. Une estimation de ρ_y dont la valeur est grande indique que les effets d'enquêteur (d_k) sur les données sont importants et l'analyse des grandes valeurs absolues de d_k permet de repérer les tâches d'enquêteur qui contribuent le plus à la variance due à l'enquêteur. On présente dans l'article des estimations de R et de ρ_y , ainsi que la distribution des d_k normalisés associés aux enquêteurs des enquêtes de suivi de 1995-1996.

3.3 Analyse des données

En vue de déterminer si l'application de la méthode proposée à l'étude de la variabilité de la déclaration de trimestre en trimestre donne de bons résultats, on a construit un ensemble de données trimestrielles recueillies en juin 1995, en septembre 1995, en décembre 1995 et en mars 1996. Aux fins de l'étude, on a éliminé aussi minutieusement que possible les variations «réelles» des stocks de trimestre en trimestre, puisque l'application satisfaisante de la méthode pour contrôler la qualité des données repose sur l'hypothèse selon laquelle les écarts entre les données de juin et celles des enquêtes trimestrielles de suivi sont des indicateurs de l'erreur de mesure. Dans la mesure où cette hypothèse est correcte, on peut appliquer la méthode pour repérer et quantifier plus facilement les sources d'erreur de mesure.

Plusieurs précautions ont été prises pour réduire au minimum les effets confusionnels des changements réels. Premièrement, on n'a considéré que deux variables observées, à savoir la superficie totale des terres de l'entreprise agricole et la superficie des terres labourables, car la valeur de ces variables est moins susceptible que d'autres de varier réellement durant l'année de référence. En outre, on a éliminé les enregistrements où figurait une modification de l'unité déclarante, indicatrice d'une modification de l'exploitation. Enfin, puisqu'on a regroupé les données des quatre trimestres (pour augmenter la taille de l'échantillon, donc

améliorer les modèles), on a également pris des mesures en vue de limiter l'influence des secteurs pour lesquels les résultats de toutes les enquêtes trimestrielles de suivi étaient cohérents, après l'enregistrement d'une incohérence initiale dans le rapport de juin. L'obtention de résultats de suivi trimestriels cohérents différents des résultats de juin pourraient signifier qu'un problème est survenu en juin. Pour résoudre cette question, on n'a retenu dans l'ensemble de données de modélisation que les données du rapport de l'enquête trimestrielle de suivi effectué le plus tôt pour les cas où des superficies identiques ont été déclarées à l'occasion des enquêtes trimestrielles de suivi subséquentes.

Des ensembles de données d'analyse distincts ont été créés pour la superficie totale des terres de l'entreprise agricole et pour la superficie des terres labourables, selon la capacité d'utiliser ces variables. Les enregistrements qui figurent dans les ensembles de données représentent chacun une réponse à une enquête trimestrielle de suivi utilisable associée à une réponse à l'enquête de juin utilisable. Pour qu'un enregistrement soit inclus dans les ensembles, la valeur de la variable étudiée doit avoir été déclarée (et non estimée ou imputée) tant en juin qu'au moment de l'enquête trimestrielle de suivi. Les ensembles résultant de données sur les quatre trimestres pour tous les états contiennent environ 7 100 enregistrements pour la superficie totale des terres de l'entreprise agricole et environ 7 000 pour la superficie des terres labourables.

3.4 Élimination des valeurs aberrantes et des points de donnée d'influence

Puisque la méthode décrite ici se fonde sur le modèle de régression par les moindres carrés, le traitement des données d'influence est très important. Les valeurs aberrantes présentent ici un dilemme particulier. Puisque nous essayons de déceler les situations problématiques, et la mesure dans laquelle on peut les attribuer aux effets différentiels des divers niveaux de la variable de groupement, nous hésitons à les écarter. Cependant, si nous n'éliminons pas certaines données aberrantes, les paramètres fondamentaux de la régression (et par conséquent, les effets de groupe) seront mal estimés. Choisir une méthode permettant de déterminer le nombre approprié de points de données d'influence à éliminer est une entreprise de taille, à laquelle beaucoup de temps a été consacré à l'occasion de la présente étude, et justifie probablement des travaux supplémentaires.

Puisque peu d'articles ont été publiés sur les méthodes de traitement des données d'influence dans les modèles mixtes (et encore moins intégrées aux logiciels), on a abordé la question de l'élimination des points de donnée d'influence dans la présente application en ignorant temporairement les effets de groupe aléatoires et en appliquant une méthode de régression linéaire. Après avoir éliminé les points de donnée d'influence repérés par régression linéaire, on a utilisé l'ensemble de données résultant pour établir le modèle avec ordonnée à l'origine aléatoire (mixte).

Plus précisément, les enregistrements dont la «valeur estimée» (par régression linéaire de la valeur de l'enquête trimestrielle de suivi en fonction de la valeur de juin) excède $3 \cdot p/n = 6/n$ (où p est le nombre de paramètres dans le modèle et n , l'effectif utilisable de l'échantillon) ont été éliminés en tant que points de donnée d'influence. Bien qu'on utilise souvent $2 \cdot p/n$ comme règle de jugement lors de l'examen des observations, le seuil de $3 \cdot p/n$ semble donner de meilleurs résultats dans le cas d'un dépistage automatisé. L'application d'une méthode automatisée est nécessaire ici puisqu'un des objectifs de l'étude consiste à mettre en oeuvre un outil de surveillance de la qualité appliqué par lot.

Après avoir éliminé les points de donnée d'influence, on a testé deux méthodes éventuelles d'élimination des données aberrantes, à savoir :

- 1.) Rajuster itérativement le modèle et, toutes les cinq itérations, éliminer les observations qui représentent plus qu'une constante, c , multipliée par la contribution moyenne des observations à la somme des carrés des résidus (représentée ici par RSS, pour *residual sum of squares*). Autrement dit, les observations dont le carré du résidu $> c \cdot \text{RSS} / n$.
- 2.) Éliminer de façon non itérative les observations qui satisfont au moins un des critères de régression linéaire définis par Belsey, Kuh, et Welsh (1980), soit :
 - a) Résidu ayant subi une transformation de Student externe > 2

- b) $|\text{Ratio de covariance}| \geq 6/n$
- c) $\text{DFFITS} > 2 \cdot \sqrt{2/n}$.

La méthode 1 semble donner de meilleurs résultats que la méthode 2, quoique, dans les deux cas, certaines observations qu'on aurait éliminées subjectivement sont retenues et certaines observations qu'on aurait retenues subjectivement sont éliminées. La méthode 2 a tendance à entraîner l'élimination d'un plus grand nombre d'observations (dans l'ensemble, environ 10 % de celles-ci), même si elle est exécutée non itérativement. Cette situation est peut-être due à l'effet de «brouillage». La méthode 1 semble mener à d'assez bonnes décisions (selon une évaluation subjective), particulièrement dans le cas des valeurs les plus élevées pour la constante c . La nature itérative de la méthode 1 semble aussi contribuer à l'élimination du masquage des valeurs aberrantes, comme en témoigne le fait que certaines observations sans influence en présence d'autres valeurs aberrantes se soient révélées très influentes durant une itération ultérieure.

On pourrait sans doute raffiner considérablement ces procédés en utilisant des méthodes de dépistage des valeurs aberrantes, comme celles examinées par Swallow et Kianifard (1996), adaptées spécialement aux situations où ces valeurs aberrantes sont multiples.

Tous les résultats publiés dans le présent article ont été obtenus en appliquant la méthode 1 d'élimination des valeurs aberrantes, en prenant $c=20$. Subjectivement, des diverses valeurs du paramètre de dispersion étudiées, cette valeur est celle qui semble donner les meilleurs résultats. Environ 8 % des enregistrements concernant la superficie totale des terres de l'entreprise agricole et environ 7 % des enregistrements concernant la superficie des terres labourables ont été éliminés à titre de valeurs aberrantes ou de points de donnée d'influence au moment de l'analyse finale.

3.5 Procédé de sélection du modèle

En s'appuyant sur la définition 3.2.1 du modèle, on a produit des modèles distincts en se servant de deux variables de groupement de base -- la tâche de l'enquêteur lors des enquêtes trimestrielles de suivi et un indicateur précisant s'il y a eu ou non changement de répondant entre l'enquête de juin et l'enquête de suivi. On a aussi envisagé un raffinement du modèle de base en vue de tenir compte de la variation de la taille de l'exploitation aux divers niveaux de la variable de groupement (c.-à-d. taille moyenne des exploitations dans la tâche d'un enquêteur). Cependant, ce raffinement n'a pas amélioré le modèle, puisque la variable calculée pour saisir cette information -- superficie moyenne en juin (totalité des terres de l'entreprise agricole ou terres labourables) pour un groupe -- s'est révélée non significative sur le plan tant statistique que pratique pour pratiquement tous les modèles. La correction proportionnelle globale, γ , paraît suffire pour tenir compte de la variation de la taille moyenne des exploitations dans chaque groupe. Cette constatation est vraie, que la variable de groupement utilisée soit la tâche de l'enquêteur ou l'indicateur de changement de répondant.

Une autre possibilité de modélisation, visant à inclure les effets aléatoires des deux variables de groupement (et leur interaction) dans le même modèle, a également été étudiée. Toutefois, cette méthode n'a pas donné de bons résultats, car le croisement de la tâche de l'enquêteur avec l'indicateur de changement de répondant produit un trop grand nombre de cellules vides ou petites. En général, on observe assez peu de changements de répondant entre l'enquête de juin et une enquête trimestrielle de suivi. Dans l'ensemble de données analysées, un tel changement ne survient que dans 25 % des cas. Par conséquent, le croisement de cette variable avec une autre n'est pas un choix de modélisation viable.

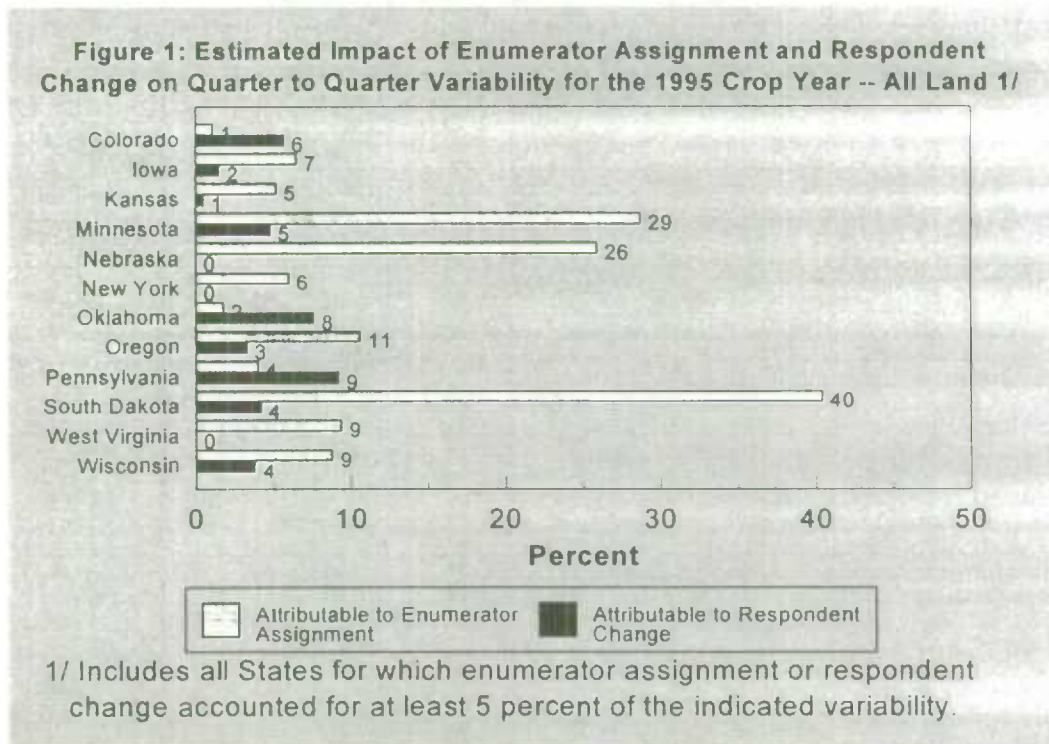
Donc, les modèles que nous avons choisis sont des modèles marginaux, dans lesquels les deux variables de groupement étudiées sont intégrées séparément. Pour éliminer tout effet confondu du modèle, tous les enregistrements comportant un changement de répondant ont été éliminés lors de l'ajustement des modèles pour lesquels la variable de groupement est la tâche de l'enquêteur.

Enfin, pour mieux refléter la variation selon l'état, les modèles ont été créés au niveau de l'état.

3.6 Résultats des modèles

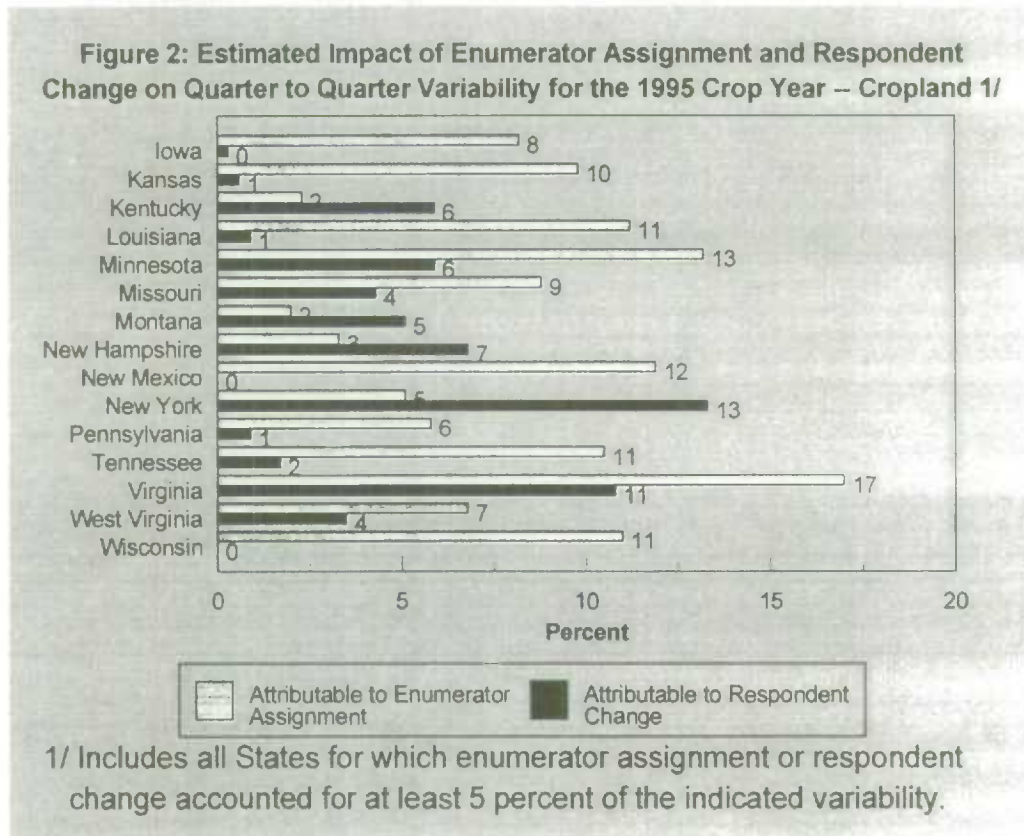
Constatation un peu surprenante, quoique, de nouveau, résultant en grande partie de la rareté de l'événement, le changement de répondant n'a habituellement que peu d'effet. Pour la plupart des états, l'effet de la tâche de l'enquêteur est plus important, et, pour certains, il est même considérable. Les figures 1 et 2 donnent une comparaison de la fraction de la variabilité du modèle attribuable à la tâche de l'enquêteur, d'une part, et au changement de répondant, d'autre part, pour la superficie totale des terres de l'entreprise agricole et pour la superficie des terres labourables, respectivement. Les pourcentages indiqués doivent toutefois être interprétés avec prudence, puisque dans quelques cas (plus précisément au Minnesota et au Nebraska pour la superficie totale des terres de l'entreprise agricole), les fortes proportions de la variabilité qui sont expliquées se rapportent à des bases très petites. Pour certains états, la variabilité globale du modèle est considérable, tandis que pour d'autres, elle est très faible.

Figure 1: Effet estimé de la tâche de l'enquêteur et du changement de répondant sur la variabilité de trimestre en trimestre pour la campagne agricole de 1995 -- Superficie totale des terres 1/



1/ Comprend tous les états pour lesquels la tâche de l'enquêteur ou le changement de répondant explique au moins 5 % de la variabilité indiquée.

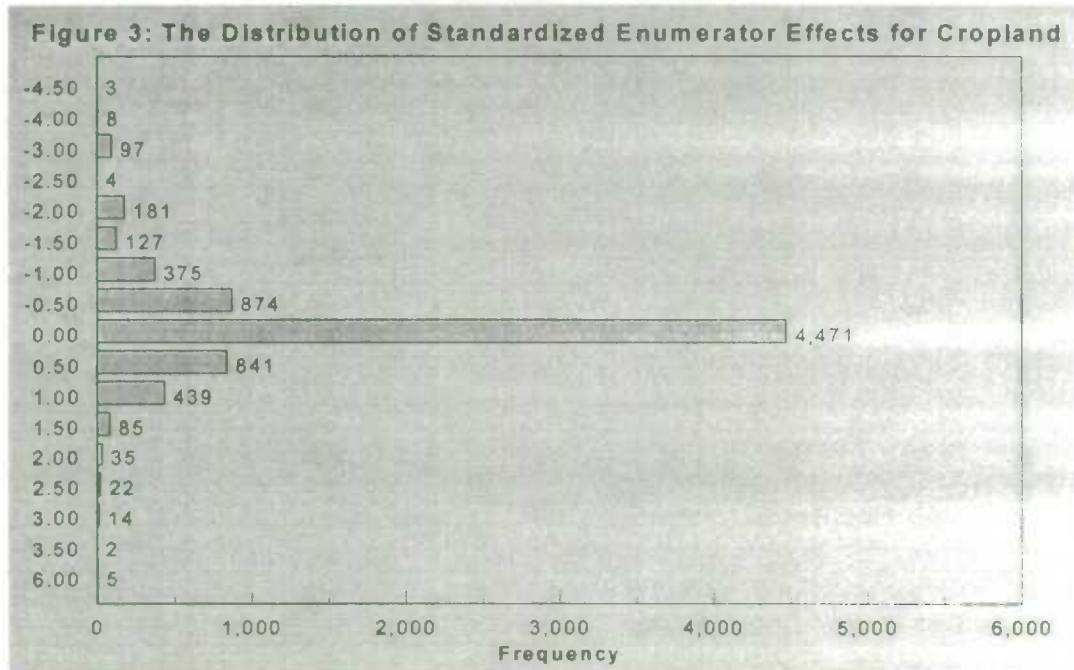
Figure 2: Effet estimé de la tâche de l'enquêteur et du changement de répondant sur la variabilité de trimestre pour la campagne agricole de 1995 -- Superficie des terres labourables 1/



1/ Comprend tous les états pour lesquels la tâche de l'enquêteur ou le changement de répondant explique au moins 5 % de la variabilité indiquée.

La figure 3 montre, pour la superficie des terres labourables, la distribution des effets estimés pour les enquêteurs, normalisée pour refléter la variabilité propre à l'état. Les queues d'une telle distribution indiquent les effets de la tâche de l'enquêteur qui sont anormalement importants et permettent de déterminer où des problèmes existent.

Figure 3: Distribution des effets normalisés d'enquêteur pour les terres labourables



Le tableau 1 résume nos estimations de la qualité des données au niveau national pour la superficie totale des terres de l'entreprise agricole et pour la superficie des terres labourables, calculées avec ou sans les valeurs aberrantes. Selon l'analyse d'«exclusion des valeurs aberrantes», les données sur les terres labourables semblent un peu moins fiables que celles sur la totalité des terres, la corrélation propre à l'enquêteur étant plus forte. En se basant sur la taille moyenne de la tâche de l'enquêteur durant les enquêtes trimestrielles visées par la présente étude ($m=17$), on pourrait s'attendre à ce qu'une corrélation propre à l'enquêteur égale à 0,008 cause une augmentation de la variance d'une estimation d'environ $(17-1)(0,008)=0,13$, soit 13 % (en appliquant l'équation 3.2.2). Les estimations de la corrélation propre à l'enquêteur varient fortement selon l'état, de sorte que l'augmentation de la variance pourrait être supérieure à cette valeur pour certains états.

Enfin, il convient de souligner l'effet important des valeurs aberrantes sur les estimations de la qualité des données. De toute évidence, les 7 à 8 % de données aberrantes dépistées dans la présente application ont une incidence marquée sur les estimations tant de la corrélation propre à l'enquêteur que de la fiabilité. Les erreurs types présentées au tableau 1 ont été calculées par la méthode du "bootstrap".

Tableau 1. Estimation du coefficient de corrélation propre à l'enquêteur et de la fiabilité

Variable observée	Scénario de dépistage	ρ_y (s.e.)	\bar{R} (s.e.)
Superficie totale des terres de l'entreprise agricole	Exclusion des valeurs aberrantes	0,004 (0,0017)	0,91 (0,005)
	Inclusion des valeurs aberrantes	0,009 (0,0043)	0,76 (0,018)
Superficie des terres labourables	Exclusion des valeurs aberrantes	0,008 (0,0027)	0,88 (0,007)
	Inclusion des valeurs aberrantes	0,010 (0,0042)	0,77 (0,017)

3.7 Conclusion

Le présent article décrit les efforts déployés en vue de se servir des données existantes pour obtenir une indication de la qualité relative des données recueillies pour diverses variables dans le cadre d'un programme permanent d'enquête, de façon à évaluer l'effet relatif des sources d'erreurs non dues à l'échantillonnage et à fournir un instrument permettant de repérer plus facilement les domaines où il est nécessaire d'offrir une formation supplémentaire.

À l'instar de toute situation où on utilise des données à des fins pour lesquelles elles ne sont pas spécialement conçues, la validité de certaines hypothèses sous-jacentes est peut-être sujette à débat. Plus précisément, on pourrait mettre en doute la validité de l'hypothèse selon laquelle la valeur de juin représente la valeur réelle, hypothèse sur laquelle se fonde en grande partie l'interprétation de nos estimations de la fiabilité et de la corrélation propre à l'enquêteur. En outre, l'hypothèse selon laquelle les groupes sont de même taille n'est pas vérifiée rigoureusement dans le cas de notre dénombrement par sondage ordinaire.

Malgré ces faiblesses, la méthode fournit des renseignements utiles. En effet, qu'on retienne ou qu'on élimine les valeurs aberrantes, elle produit des estimations de la qualité de l'enquête comparables à celles obtenues antérieurement, de façon beaucoup plus coûteuse, en effectuant des enquêtes par réinterview. Il reste à déterminer quelles estimations sont les meilleures et comment devrait être exécuté le dépistage des valeurs aberrantes (ou, tout simplement, s'il doit être exécuté). Le fait que 7 à 8 % de données aient un effet significatif sur les estimations calculées indique toutefois la présence d'observations ayant une forte influence dont il conviendrait de vérifier la validité. L'effet important de ces observations soulèvent également la question de savoir si un modèle plus robuste, sans dépistage des valeurs aberrantes, fournit des résultats encore meilleurs. Cette question fera l'objet d'une étude ultérieure.

À cause de la taille inégale des groupes et de l'amalgame des erreurs, les tâches d'enquêteur problématique selon l'analyse ne sont pas nécessairement indicatrices d'un mauvais dénombrement et témoignent parfois de la combinaison d'une tâche de petite taille et d'une erreur grave de saisie de données. Toutefois, quelle que soit la raison de leur dépistage, les groupes pour lesquels on note un effet de groupe important en valeur absolue sont généralement ceux qu'il convient d'examiner.

Enfin, le programme informatique utilisé pour effectuer l'analyse a été conçu avec l'intention de l'utiliser éventuellement comme outil permanent destiné à faciliter la tâche des statisticiens de notre bureau de la statistique chargé de recueillir les données. Il produit des listes récapitulatives des valeurs aberrantes et des points de donnée d'influence éliminés lors de la modélisation, et de tous les échantillons pour lesquels l'effet de la tâche de l'enquêteur est statistiquement significatif. Les résultats de l'étude indiquent que ce programme pourrait être utilisé au niveau opérationnel.

4. BIBLIOGRAPHIE

- Belsley, D.A., Kuh, E., et Welsch, R.E. (1980), "Regression Diagnostics", New York: John Wiley & Sons, Inc.
- Biemer, P. et Atkinson, D. (1995). "An Integrated Approach for Estimating Measurement Error Bias and Variance in Two-Phase Samples," *Proceedings of the 1995 Annual Research Conference*, 355-357.
- Biemer, P. et Atkinson, D. (1995). "Estimating Measurement Error Bias and Variance in Two-Phase Samples," *American Statistical Association - Proceedings of Section on Survey Research Methods*, Volume II, 775-780.
- Biemer, P. et S.L. Stokes (1991). "Approaches to the Modeling of Measurement Errors." In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, S. Sudman (Eds.) *Measurement Errors in Surveys*. New York: John Wiley & Sons, 487-516.
- Groves, R.M. (1989), "Survey Errors and Survey Costs", John Wiley & Sons, N.Y.

Swallow, W. et Kianifard, F. (1996). "Using Robust Scale Estimates in Detecting Multiple Outliers in Linear Regression," *Biometrics*, **52**, 545-555.

Wright, R.L. (1983). "Finite Population Sampling with Multivariate Auxiliary Information," *Journal of the American Statistical Association*, **78**, 879-884.

SESSION 8

DONNÉES ADMINISTRATIVES

QUELS SONT LES IMPACTS SUR LA QUALITÉ DES DONNÉES LORSQUE DES DONNÉES D'ENQUÊTE SUR LE REVENU SONT COMBINÉES À DES DONNÉES ADMINISTRATIVES?

Sylvie Michaud et Michel Latouche¹

RÉSUMÉ

En général, l'erreur de mesure pose des problèmes. Beaucoup de travaux ont été effectués à ce sujet, à la fois pour mesurer l'erreur et pour la compenser. Il a été démontré que l'erreur de mesure peut être plus problématique dans le cadre d'une enquête longitudinale, particulièrement si les données sont utilisées pour effectuer des régressions. L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui vise à déterminer les répercussions qu'ont sur le revenu les changements touchant l'activité sur le marché du travail et les caractéristiques familiales. Afin de réduire le fardeau de réponse et d'améliorer la qualité des données, on offre un choix aux répondants dans le cadre de l'enquête, à savoir répondre aux questions sur le revenu ou autoriser les responsables de l'EDTR à avoir accès à leurs dossiers administratifs. Le présent document vise à quantifier les répercussions de cette approche mixte sur l'erreur de réponse, et plus particulièrement sur les mesures du changement.

MOTS CLÉS: Données longitudinales; erreurs de réponse; sous-déclaration; taux de réponse; tendances.

1. INTRODUCTION

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui mesure les répercussions qu'ont sur le revenu les changements touchant l'activité sur le marché du travail ou la situation familiale. Les personnes comprises dans un panel donné demeurent dans l'échantillon pendant six ans et sont interviewées deux fois par année. En janvier, on recueille des données sur le travail, tandis qu'en mai, on recueille des données sur le revenu. L'interview sur le revenu est effectuée en mai parce que les Canadiens doivent produire leur déclaration de revenu avant la fin d'avril et que l'on croit de façon générale qu'ils sont plus susceptibles de fournir des données précises sur leur revenu à ce moment-là. En mai 1995, dans le cadre d'un effort en vue de réduire le fardeau de réponse, on a offert aux répondants de l'EDTR de répondre aux questions sur le revenu ou d'autoriser les responsables de l'enquête à obtenir des données sur leur revenu à partir de leur déclaration de revenu. Cette offre leur est faite chaque année et, après trois ans de collecte, plus de 75 % des répondants se prévalent de la deuxième option. Toutefois, l'intégration des données d'enquête et des données fiscales ne se fait pas sans problème. Les définitions ne sont pas toujours compatibles et il existe des problèmes de couplage. D'un autre côté, il faut toutefois tenir compte des problèmes de qualité des données qui découlent généralement des enquêtes sur le revenu (par exemple la sous-déclaration de certaines sources de revenu) et de la nécessité de procéder à une imputation. Le présent document donne un aperçu des diverses sources d'erreurs que présente cette méthode et fait état de certaines des répercussions que comporte cette approche mixte. La recherche a été axée sur les microcomparaisons et on a tenté de quantifier les répercussions sur les mesures du changement.

¹S. Michaud et M. Latouche, Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa, Ontario, K1A 0T6.

2. PLAN D'ÉCHANTILLONNAGE DE L'EDTR

L'échantillon de l'EDTR est choisi à partir d'un plan d'échantillonnage à plusieurs degrés. Les répondants sélectionnés pour faire partie de l'échantillon de l'EDTR ont préalablement participé à l'Enquête sur la population active (EPA) pendant six mois. On les interviewe deux fois par année pendant six ans. Une première interview, qui se tient en janvier, sert à recueillir des données détaillées sur le travail. Elle permet en outre d'avoir un aperçu des changements qui touchent la composition de la famille, et des dates où ces changements se sont produits. La deuxième interview, qui se tient en mai, sert à recueillir des données détaillées sur le revenu, selon 24 catégories. Dans le cadre de l'enquête, on recueille aussi des données sur l'impôt sur le revenu versé, afin d'évaluer le revenu après impôt. Des données sur le revenu sont recueillies pour chaque membre du ménage âgé de 16 ans et plus et sont agrégées au niveau de la famille, afin de déterminer les mesures de faibles revenus.

La collecte de données sur le revenu n'est pas nouvelle à Statistique Canada. Dans le cadre de l'Enquête sur les finances des consommateurs (EFC), on recueille des données annuelles sur le revenu depuis les 30 dernières années, et les questions sur le revenu de l'EDTR sont identiques à celles de l'EFC. L'expérience acquise dans le cadre de cette dernière est très utile pour l'EDTR.

En général, les enquêtes sur le revenu obtiennent un taux de réponse plus faible que les autres enquêtes. Alors qu'une enquête qui ne porte pas sur le revenu, comme l'Enquête sur la population active, obtient généralement un taux de réponse de 95 %, le taux de réponse de l'Enquête sur les finances des consommateurs n'est que de 80 %. Quant à l'interview sur le revenu de l'EDTR, elle obtient un taux de réponse de 76 %. Les données ont aussi été couplées avec celles d'autres sources pour en évaluer la qualité. À partir de ces comparaisons, on a déterminé qu'il y a sous-déclaration de certaines sources de revenu, par exemple les prestations d'assurance-chômage, les prestations d'aide sociale, ainsi que les intérêts et les dividendes[5].

Les données fiscales ont aussi été utilisées plus récemment comme source de données sur le revenu. De façon plus particulière, le fichier de données administratives longitudinales (DAL) est un fichier longitudinal fondé sur les données fiscales [7]. Un échantillon de 10 % des répondants à l'enquête a été choisi au hasard et les familles ont été rebâties, selon les renseignements fournis dans les déclarations de revenu (les conjoints et les enfants ont été créés à partir d'autres champs de la déclaration de revenu). Le fichier DAL ne correspond pas exactement aux autres sources de données administratives. On ne peut créer que des familles de recensement (c.-à-d. père-mère-enfants), et on a tendance à surestimer le nombre de familles ne comptant qu'une personne. Il y a aussi sous-représentation de certains groupes d'âge (particulièrement les personnes plus âgées) et de faible revenu. Récemment, toutefois, avec l'avènement des crédits d'impôt, on obtient à partir du système fiscal une meilleure couverture de la population de l'univers. Le problème de la reconstruction des familles subsiste toujours. Par ailleurs, la qualité des données sur le revenu provenant de sources fiscales est jugée supérieure à celle des données d'enquête.

Dans le cadre de l'EDTR, on utilise les deux approches, afin de maximiser les taux de réponse et la qualité des données. Cette façon de faire comporte toutefois des problèmes.

3. SOURCES THÉORIQUES DES ERREURS DÉCOULANT DE L'UTILISATION DE DONNÉES D'ENQUÊTE ET DE DONNÉES FISCALES

Le principe de l'utilisation de données administratives comme sources de données sur le revenu comporte un certain nombre d'enjeux. Par exemple, le moment de la production de ces données peut avoir des répercussions sur les dates cibles de l'enquête. Le couplage avec les fichiers administratifs avec ou sans identificateur unique peut aussi poser certains problèmes, et si l'on tente d'évaluer les répercussions globales d'une stratégie mixte, on doit tenir compte de ces éléments. Toutefois, dans le cas présent, la discussion se limitera aux questions de couplage des deux sources, du point de vue de la qualité des données. Une discussion générale de l'utilisation des dossiers fiscaux dans le cadre de l'EDTR figure dans l'ouvrage [1] de la bibliographie.

Afin d'assurer une mesure globale de la qualité, les enquêtes doivent permettre de calculer l'erreur quadratique moyenne, c'est-à-dire la somme de la variance d'une variable donnée et le carré du biais. Cette opération est généralement ardue parce que le biais peut être très difficile à mesurer. Le tableau 1 vise à déterminer les avantages ou inconvénients possibles de chacune des méthodes (données d'enquêtes et données fiscales) ainsi qu'à préciser sur quelle composante de l'erreur elles peuvent avoir des répercussions.

La couverture n'est touchée que par l'utilisation des données fiscales. Le champ des dossiers fiscaux s'est élargi au fil des ans, et il comprend maintenant 94 % de la population âgée de 20 ans et plus. Si la population des personnes qui ne produisent pas de déclaration de revenu diffère de celle constituée des personnes qui en produisent une, cela peut entraîner un biais dans les données. Les étudiants, par exemple, sont un groupe susceptible d'être sous-représenté chez les personnes qui produisent une déclaration de revenu; le fait qu'ils sont généralement associés à un plus faible revenu et qu'il peut exister une différence entre les étudiants qui produisent une déclaration de revenu et ceux qui n'en produisent pas peut créer un biais.

Tableau 1. Comparaison de la collecte au moyen de l'enquête et de l'utilisation des données fiscales.

	Enquête seulement	Données fiscales seulement
couverture de la population (biais)		↘ personnes qui produisent une déclaration de revenu uniquement
taux de réponse (total) (variance) (biais)	↘ nature délicate du sujet fardeau de réponse dépistage	↗ toutes les personnes qui produisent une déclaration de revenu sont des «répondants»
		↘ aucun couplage ou couplage inapproprié
erreur de réponse (biais)	↘ ↘ sous-déclaration de certaines sources de revenu (assurance-chômage, intérêts...)	↘ sous-déclaration de certaines sources de revenu (économie souterraine)
	↘ arrondissement déclaration par personne interposée	
		↘ sources de revenu non imposable

	Enquête seulement	Données fiscales seulement
cohérence des séries chronologiques (biais)	↘ erreur de réponse et incohérences longitudinales	
		↘ incohérences possibles quant aux définitions des catégories de revenu

↗ aspect positif ↘ aspect négatif

La non-réponse peut entraîner des problèmes, tant du point de vue de la variance que du biais. Le revenu est un sujet délicat pour certains répondants et les questions s'y rapportant tendent à obtenir un taux de réponse «plus faible» dans le cadre d'une enquête. Le fait que l'EDTR soit une enquête longitudinale a aussi des répercussions sur le taux de réponse. Au fil des ans, des personnes déménagent, et l'impossibilité de les dépister a aussi pour effet de diminuer le taux de réponse. Le niveau de différence entre les non-répondants et les répondants déterminera l'amplitude du biais. L'utilisation des données fiscales devraient compenser pour certains de ces problèmes en théorie. Lorsqu'une personne produit une déclaration de revenu, il devrait être possible de retracer son dossier fiscal, ce qui devrait augmenter le taux de réponse. Toutefois, dans le cadre de l'EDTR, on ne recueille pas de numéros d'assurance sociale, lesquels constituent un lien unique avec le dossier fiscal. D'autres champs, qui seront décrits à la section 4, sont utilisés dans le cadre d'une méthode de couplage statistique pour établir un lien entre les personnes qui appartiennent à l'échantillon de l'EDTR et les dossiers fiscaux les concernant. Certaines mesures de contrôle de la qualité sont utilisées pour améliorer la qualité du couplage, mais il existe toujours une possibilité de couplage erroné ou de non-couplage, même dans le cas d'une personne qui produit une déclaration de revenu. Cela aussi a pour effet de diminuer le taux de réponse.

L'erreur de réponse a été étudiée dans le cas de certaines variables du revenu pour lesquelles il existait une source externe en vue de valider les résultats et d'évaluer les biais possibles. Certaines études ont fait ressortir qu'il y avait sous-déclaration de certaines sources de revenu pour les données recueillies dans le cadre d'enquêtes. Pour l'EFC, par exemple, on ne saisit qu'environ 80 % des prestations d'assurance-chômage, comparativement à 94 % grâce au système fiscal. Les revenus de placements sont aussi sujets à la sous-déclaration. Cela crée un biais dans les résultats. Par ailleurs, on pense de façon générale que les données fiscales sont sujettes à la sous-déclaration dans le cas de certaines sources de revenu qui sont liées à l'économie souterraine. Toutefois, étant donné que dans le cadre de l'EDTR on demande aux répondants de consulter leurs dossiers fiscaux pour fournir des renseignements sur leur revenu, et étant donné qu'il n'est pas évident que ces répondants déclareraient ce genre de revenus dans le cadre d'une enquête, on peut conclure que les données fiscales peuvent aussi donner lieu à un biais, ce biais n'étant toutefois pas aussi prononcé que celui des données d'enquête.

L'arrondissement des montants de revenu déclarés dans le cadre d'une enquête constitue une deuxième source d'erreur de réponse. L'arrondissement du revenu total déclaré pose un problème, que vient aggraver l'arrondissement à partir de diverses sources de revenu, le revenu total correspondant alors à la somme de ces sources.

Une troisième source d'erreur de réponse, qui touche les données fiscales, est l'absence de certaines sources de revenu dans le formulaire de déclaration de revenu. Certains revenus non imposables sont déclarés, mais ils se limitent à ceux qui doivent l'être pour le calcul des crédits d'impôt. Les revenus provenant des gains de

loterie et d'héritages ne figurent pas dans la déclaration de revenu, mais sont compris dans la collecte effectuée dans le cadre de l'enquête.

Les changements quant aux définitions et aux concepts fiscaux constituent une quatrième source d'erreur de réponse. Les séries chronologiques peuvent être affectées par les changements apportés à la réglementation touchant l'impôt sur le revenu.

Comme on peut le constater, les deux sources de données comportent des problèmes. La présente étude vise à déterminer les répercussions de la collecte mixte effectuée dans le cadre de l'EDTR sur la qualité des données. Étant donné la nature longitudinale de l'EDTR, les mesures du changement sont importantes. Les erreurs de réponse posent davantage de problèmes dans le cadre d'une enquête longitudinale que d'une enquête transversale, étant donné qu'on s'attend généralement à ce que la corrélation entre les mesures répétées soit plus grande que la corrélation entre les erreurs de réponse. Étant donné les arrondissements possibles et la sous-déclaration des erreurs, on s'attend à ce que les données de sources administratives aient un taux d'erreur de réponse inférieur à celui des données d'enquête.

De façon plus particulière, partons de l'hypothèse qu'on veuille quantifier une variable X (revenu), mais qu'on mesure $x = X + u$ en réalité, u correspondant à l'erreur de réponse. Dans une régression, où on désire obtenir: $Y = X\beta + \epsilon$, on mesure en réalité: $y = x\beta' + \epsilon$ où le biais de β' tend vers zéro, avec les hypothèses habituelles d'erreurs indépendantes et normales. Si on voulait analyser le changement $\Delta Y = Y_{t+1} - Y_t$ par régression, on constate dans [1] que le biais dans l'équation du changement est plus grand que le niveau, ce qui s'exprime par la formule mathématique

$$\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < \frac{\sigma_{\Delta x}^2}{\sigma_{\Delta x}^2 + \sigma_{\Delta u}^2}$$

4. DÉTERMINATION EMPIRIQUE DES SOURCES D'ERREURS

Pour la première année de collecte des données de l'EDTR, on n'a pas utilisé les déclarations de revenu et on a recueilli les données sur le revenu directement auprès des répondants. Afin d'assurer la qualité des données, on a procédé à une comparaison entre les données d'enquête et les données fiscales. Cela a permis de déterminer et de quantifier les sources d'erreurs.

Les données fiscales sont le résultat d'un couplage, mais étant donné que l'on ne demande pas aux répondants leur numéro d'assurance sociale (NAS), le couplage se fait au moyen d'une méthode statistique. Les dossiers ont d'abord fait l'objet d'un couplage direct à partir du nom, du code postal, de la date de naissance, du sexe et de l'état matrimonial. Grâce à cette méthode, on a couplé 50 % des dossiers. Les dossiers qui n'ont pas été couplés ont fait l'objet d'un couplage statistique (tenant compte des valeurs manquantes ou des écarts pour un ou plusieurs des éléments de couplage). Cela a donné lieu à un couplage global de 85 %. L'étude a mis l'accent sur le taux de réponse, la couverture, le couplage et les erreurs de réponse. Une attention spéciale a été accordée aux répercussions sur les tendances annuelles.

4.1 Taux de réponse et biais possibles

Le fichier de l'échantillon de l'EDTR a été couplé aux dossiers fiscaux de 1993, à partir de méthodes de couplage statistique et direct. Le tableau 2 présente la distribution de l'échantillon, selon le taux de réponse à l'interview sur le revenu et les résultats du couplage avec les données fiscales.

Tableau 2. Taux de réponse en fonction du couplage avec les données fiscales.

	Aucun couplage avec les données fiscales	Couplage avec les données fiscales	Total
Répondants	3 605	20 651	24 256 (76 %)
Non-répondants	1 774	5 709	7 483 (24 %)
Total	5 379 (17 %)	26 360 (83 %)	31 739

Si toutes les personnes dont les réponses ont fait l'objet d'un couplage avec les dossiers fiscaux avaient accepté cette façon de faire, on aurait assisté dans les faits à une augmentation du taux de réponse. Toutefois, seulement 75 % des répondants ont approuvé l'utilisation de leurs données administratives. Par ailleurs, dans le cadre de l'EDTR, on tente aussi de recueillir des données sur le revenu des personnes qui disent ne pas remplir de déclaration de revenu. Dans l'ensemble, deux groupes de personnes pourraient avoir eu des répercussions sur les taux de réponse. D'abord, les non-répondants à l'interview sur le revenu ayant autorisé l'accès à leurs données fiscales et ayant fait l'objet d'un couplage auraient eu un effet positif sur le taux de réponse. Viennent ensuite, les personnes ayant autorisé l'accès à leur dossier fiscal, mais n'ayant pas fait l'objet d'un couplage, auraient eu un effet négatif. Chacun des deux groupes comportait environ 1 700 personnes. Cela signifie que le taux de réponse est demeuré le même avec la méthode mixte.

Toutefois, en raison des biais possibles découlant de différences entre les répondants ayant fait l'objet d'un couplage et ceux n'ayant pas fait l'objet d'un couplage, on a comparé ces deux groupes, à partir des données sur le revenu recueillies au cours de la première année de l'enquête. Trois sous-groupes comportaient des différences marquées: les célibataires âgés de 15 à 19 ans, les célibataires âgés de 20 à 24 ans et les femmes mariées âgées de 45 ans et plus. Chez ces groupes, on notait un pourcentage élevé de dossiers non couplés et, de façon générale, les revenus des personnes ayant fait l'objet d'un couplage et de celles n'ayant pas fait l'objet d'un couplage étaient différents (ces dernières avaient un revenu inférieur). Par la suite, on a comparé les groupes de personnes n'ayant pas fait l'objet d'un couplage, selon que ces personnes avaient donné leur autorisation ou non. Cinq grandes catégories ont été utilisées pour effectuer ces comparaisons : les revenus d'emploi (salaire et revenu d'emploi autonome), les revenus de placements (revenus de placement imposables, y compris intérêts et dividendes), les paiements de transfert gouvernementaux (assurance-chômage, aide sociale, prestation fiscale pour enfants, sécurité de la vieillesse, Régime de pension du Canada, indemnisation des accidents du travail et crédits pour la taxe sur les produits et services (TPS)), et le revenu total. Les comparaisons ont été effectuées pour un sous-ensemble de dossiers étiquetés répondants «acceptables». On a procédé ainsi pour écarter les effets possibles de l'imputation. Les résultats figurent au tableau 3. Une tendance similaire s'est dégagée pour toutes les catégories de revenu. Cela laisse supposer que si l'on définissait un ensemble approprié de répondants, un modèle d'imputation relativement valide pourrait être établi pour les personnes n'ayant pas fait l'objet d'un couplage, étant donné qu'il ne semble pas y avoir de différence entre celles qui ont donné leur autorisation et celles qui ne l'ont pas fait.

Tableau 3. Comparaisons du revenu total des répondants «acceptables» (à partir des données d'enquête) pour les répondants qui ont autorisé l'utilisation de leurs dossiers administratifs et pour ceux qui ne l'ont pas fait.

	«répondants acceptables» qui ont donné leur autorisation			«répondants acceptables» qui n'ont pas donné leur autorisation		
	n	moyenne	médiane	n	moyenne	médiane
célibataires âgés de 15 à 19 ans	865	2 624 \$	1 500 \$	727	2 458 \$	1 000 \$
célibataires âgés de 20 à 24 ans	606	10 987 \$	8 800 \$	470	9 623 \$	7 188 \$
femmes mariées âgées de 45 ans et plus	1 599	12 771 \$	7 677 \$	1 183	13 573 \$	8 160 \$
autres	7 857	25 657 \$	20 000 \$	5 509	26 667 \$	21 567 \$

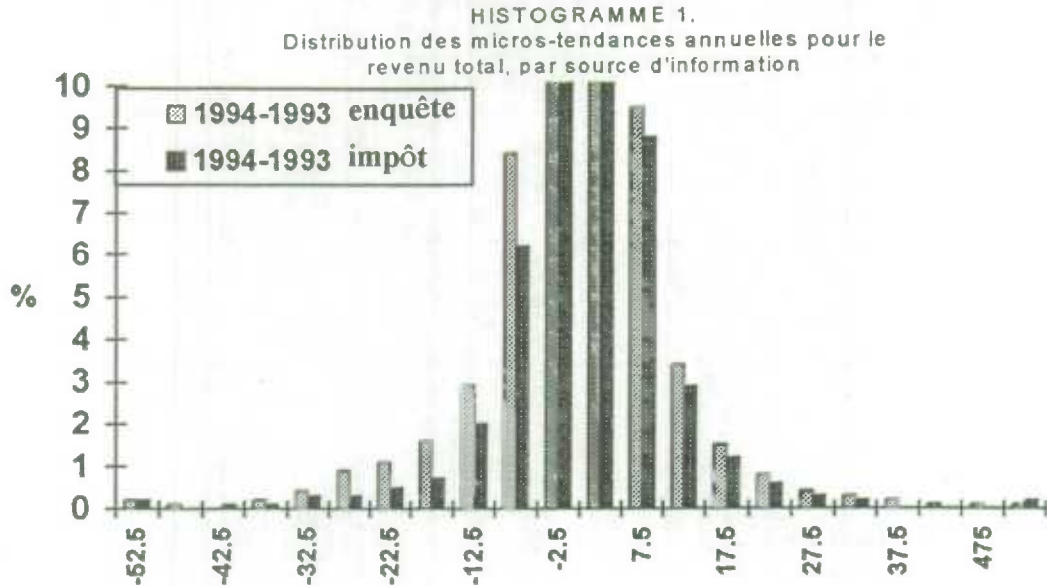
4.2 Erreur de réponse

Des comparaisons ont été effectuées pour un certain nombre de catégories de revenu à partir des données d'enquête et des données fiscales. La référence [5] parle d'une comparaison entre des données d'enquête et des données fiscales pour les répondants «acceptables», en vue de déceler les différences possibles quant aux définitions des catégories de revenu. Leurs conclusions laissent supposer qu'il existait des différences dans le cas des revenus du travail autonome et de l'aide sociale. Il n'existait pas de différence entre le revenu moyen et le revenu médian pour les salaires et les prestations d'assurance-chômage (a.-c.). Toutefois, on notait encore une certaine sous-déclaration des prestations d'assurance-chômage. La difficulté de cette méthode a consisté à déterminer ce qui était juste. De façon plus particulière, lorsqu'il y avait des différences quant à la déclaration de revenus d'emploi autonome, il était difficile de déterminer s'il s'agissait de revenus de l'économie souterraine, lesquels n'auraient pas figuré dans les déclarations d'impôt, de revenus déclarés provenant d'une autre source ou encore de revenus mal déclarés. Étant donné que l'EDTR est axée principalement sur l'analyse longitudinale, il a été décidé d'étudier les différences quant aux mesures du changement et de tenter de concilier les microdifférences des dossiers comportant deux années de données. Cela permet en outre d'étudier l'erreur de réponse en tenant compte de cet aspect longitudinal.

On ne disposait de deux années de données d'enquête et de données fiscales que pour un sous-ensemble seulement des répondants de l'EDTR. Cela a limité l'étude à un sous-échantillon de 4 274 répondants. Ce sous-échantillon n'est pas très représentatif de l'échantillon global; il comporte un pourcentage légèrement plus élevé de personnes âgées de 65 ans et plus et une représentation plus faible du groupe des très jeunes, âgés de 16 à 19 ans. Les écarts n'ont pas été jugés suffisamment importants pour invalider l'étude. Dans ce sous-ensemble, 86 % des dossiers avaient été obtenus grâce au couplage direct, ce qui fait que l'on a conservé uniquement ce sous-ensemble, pour supprimer les répercussions possibles des dossiers mal couplés. Les comparaisons se sont limitées à ce sous-ensemble de 3 670 dossiers. Toutefois, on a supprimé 600 autres dossiers, dont 400 comportaient une non-réponse partielle pour la deuxième année, et la majeure partie des 200 autres, aucun revenu déclaré pour une année ou les deux.

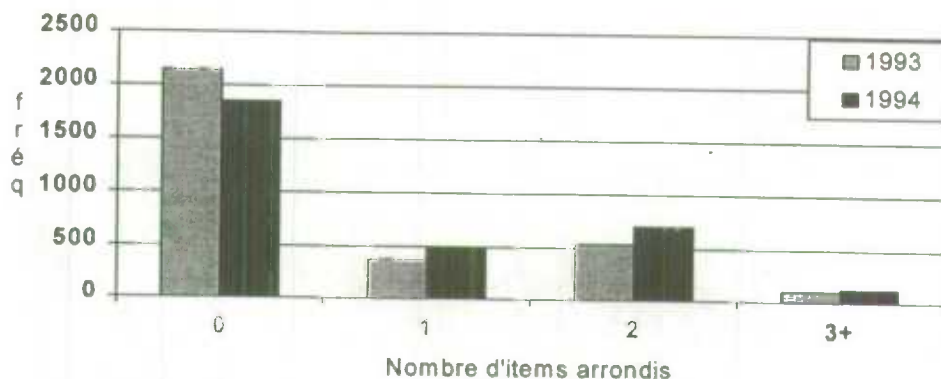
On a accordé une attention particulière aux mesures du changement quant au revenu total entre les deux années, à partir à la fois des données d'enquête et des données fiscales. Pour chaque personne, un changement de revenu total ou micro-tendance a été calculé pour les données d'enquête et les données fiscales. L'histogramme 1 montre la distribution des changements. L'axe des ordonnées a été limité à 10 % pour avoir

un meilleur aperçu des queues de la distribution (l'axe aurait dû atteindre près de 30 %). Le fait d'utiliser les données fiscales pour calculer le changement moyen a eu pour résultat d'augmenter le revenu de 498\$, tandis



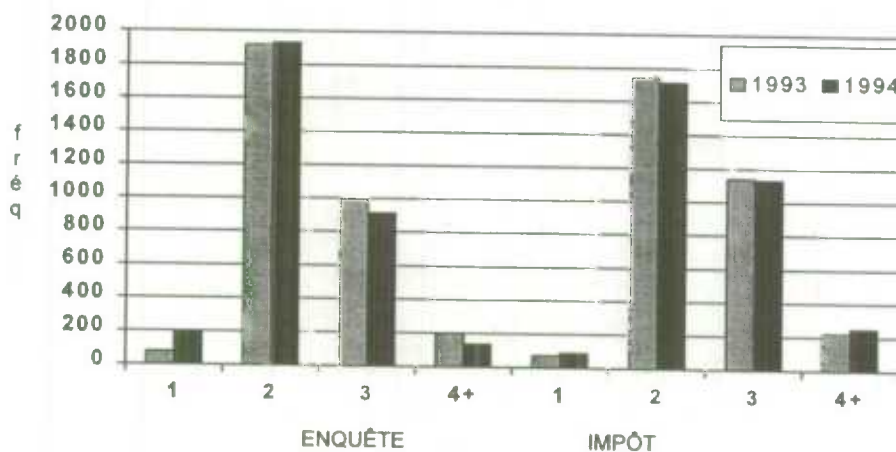
que le recours aux données d'enquête a eu pour effet de le diminuer de 3\$ (l'écart était significatif au niveau de 1 %). Il semble en outre y avoir plus de variabilité dans les mesures du changement à partir de l'enquête. Puis, on a comparé les modèles de déclaration. Un examen initial des données a fait ressortir deux comportements différents, selon qu'une personne fournissait des montants approximatifs (ce que l'on a décelé en examinant les sources de revenu ayant fait l'objet d'un arrondissement) ou des montants exacts. L'étude a donc mis l'accent sur le niveau d'arrondissement. L'étude s'est limitée à l'arrondissement de dix catégories de revenu, les autres catégories de revenu n'étant pas déclarées de façon similaire dans les données d'enquête et les données fiscales. On a déterminé que les montants arrondis étaient ceux dont les deux derniers chiffres étaient de zéro dans les données d'enquête, mais pas dans les données fiscales. L'histogramme 2 montre la distribution des répondants selon le nombre d'éléments de revenu arrondis dans les données d'enquête pour les années de référence 1993 et 1994. Il arrive fréquemment que les montants de revenu soient arrondis. En effet, seulement 1 530 dossiers, ce qui représente 47,5 % des personnes, ne comportaient pas de montants arrondis de revenu pour les deux années. Il semble en outre que le niveau d'arrondissement ait augmenté pour la deuxième année de collecte.

HISTOGRAMME 2
Distribution du nombre d'items arrondis par individus
par an basé sur 10 items



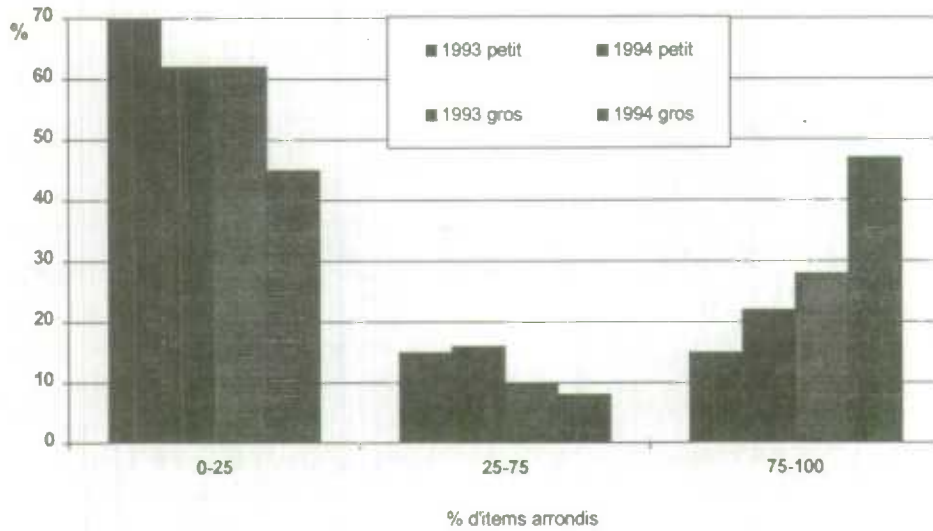
L'histogramme 3 montre la distribution des répondants selon le nombre d'éléments non nuls déclarés dans le cadre de l'enquête. Il est intéressant de noter que le nombre moyen d'éléments déclarés dans le cadre de l'enquête a diminué légèrement pour la deuxième année d'enquête, par rapport à la première année, mais cela ne ressort pas des données fiscales déclarées pour ces mêmes éléments.

HISTOGRAMME 3
Distribution des items non-nuls par individu
par année et par source



Il était intéressant de déterminer si l'arrondissement a été différent pour les divers groupes de répondants. On a comparé le comportement en matière d'arrondissement au fil des ans pour divers groupes de revenu. Comme le montre l'histogramme 4, le revenu semble avoir des répercussions sur l'arrondissement, c'est-à-dire que les personnes qui ont un revenu élevé ont tendance à l'arrondir davantage que celles qui ont un revenu faible.

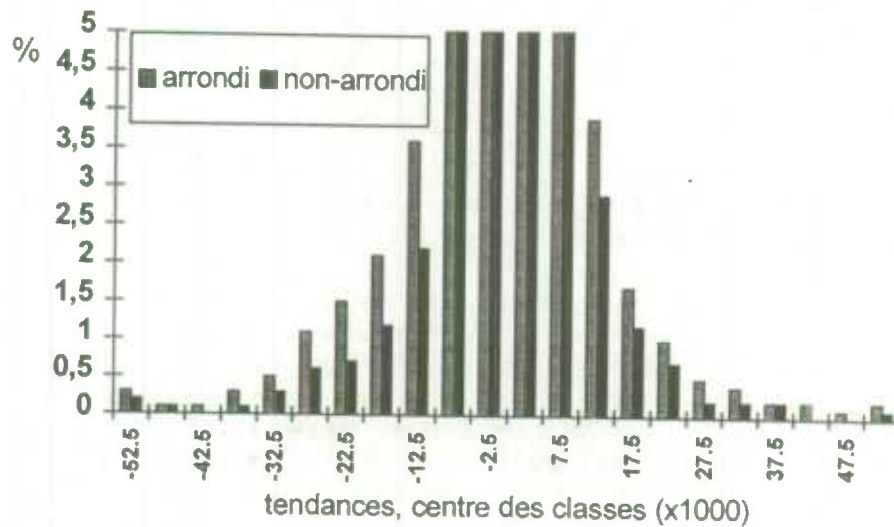
HISTOGRAMME 4
 Distribution des individus par proportions d'items arrondis
 années de référence 1993 et 1994
 pour les répondants avec un petit et un gros revenu total



Les personnes âgées, quant à elles, semblent moins susceptibles d'arrondir les montants qu'elles fournissent.

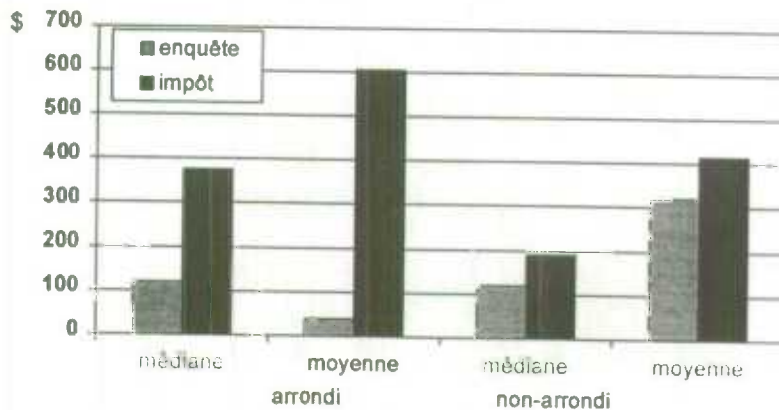
Encore une fois, on a comparé le revenu total. Les mesures du changement ou des microtendances ont fait l'objet de comparaisons à partir des données d'enquête, après division des répondants en deux groupes : les répondants ayant arrondi leur revenu déclaré et les répondants ne l'ayant pas arrondi. Les résultats figurent à l'histogramme 5. Un répondant a été mis dans le groupe ayant fourni des revenus arrondis s'il avait arrondi au moins un des éléments déclarés. Il est intéressant de noter qu'une proportion importante de la variabilité des mesures du changement observée à l'histogramme 1 a trait au groupe de répondants qui ont arrondi leur revenu pour au moins une source.

HISTOGRAMME 5
Distribution des micros différence (1994 - 1993)
pour les groupes "arrondi" et "non-arrondi"



Pour confirmer cette hypothèse, les mesures du changement ont été comparées entre les données d'enquête et les données fiscales pour le groupe de répondants qui n'avaient pas arrondi les montants de revenus déclarés.

HISTOGRAMME 6
Mesures de centralité des micro-tendances
pour les groupes "arrondi" et "non-arrondi"



L'histogramme 6 indique la microtendance moyenne et médiane séparément pour le groupe aux montants «arrondis» et celui aux montants «non arrondis». Dans le cas des personnes qui n'ont pas arrondi leur revenu, les différences n'étaient pas si importantes. Comme on pouvait s'y attendre, les écarts les plus grands ont été enregistrés pour le groupe des personnes ayant arrondi leur revenu.

5. CONCLUSIONS

Nous n'avons fait ici qu'un survol de la question de l'erreur de réponse. Lorsque certaines des différences les plus importantes ont été examinées par des spécialistes, elles ont été attribuées à une erreur de réponse dans le cadre de l'enquête pour approximativement 80 % des cas. Environ 10 % des différences ont été attribuées à une «erreur» dans les données fiscales (un élément non imposable était absent pour une des deux années ou il semblait y avoir une erreur dans le champ de la déclaration de revenu). Enfin, les autres 10 % de différences n'ont pas pu être expliquées.

On est arrivé à certaines conclusions intéressantes, par exemple : environ 30 % des gens étudiés ont fourni exactement les mêmes montants (au dollar près) dans le cadre de l'enquête et dans les dossiers fiscaux, pour au moins une année de données. Pour le reste des répondants, une erreur de réponse a été notée dans le cadre de l'enquête ou dans les dossiers fiscaux, ou encore dans les deux. Les sources de données comportent toutes deux leurs limites; celles des données fiscales ont trait aux personnes qui ne remplissent pas de déclaration de revenu et à la sous-déclaration des revenus non imposables, tandis que les données d'enquête semblent donner lieu à des erreurs de réponse. L'erreur de réponse relative aux données d'enquête semble aussi augmenter avec le temps. À partir des résultats observés, et même si les données fiscales sont sujettes à erreur, le recours à ces dernières dans le cadre de cette approche mixte entraînera probablement une amélioration de la qualité des données sur le revenu, plus particulièrement du fait de la nature longitudinale de l'enquête. Il subsiste des questions à examiner. En effet, les conclusions générales ne semblent pas s'appliquer aussi bien aux travailleurs autonomes. Ce groupe devrait donc faire l'objet d'une analyse plus poussée. De même, l'étude devrait être plus précise du point de vue de la déclaration des revenus selon la source. Les répercussions des revenus non imposables sur la mesure du changement devraient aussi être évaluées de façon plus détaillée. Enfin, un certain nombre des techniques ont été proposées pour tenir compte de l'erreur de réponse dans les ouvrages [3] et [6] de la bibliographie. Ces techniques devraient être appliquées et mises à l'essai pour vérifier si elles peuvent contribuer à améliorer la qualité des mesures du revenu.

Les auteurs voudraient souligner la contribution de Chantal Grondin, Martin Renaud, Carole Janelle et Elaine Fournier pour la préparation de l'étude.

6. BIBLIOGRAPHIE

Bound, J., Brown, C., Duncan, G. et Willard, R. (1991), «Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Validation Study Evidence», *Panel Data and Labor Market Studies*, J. Hartog, G. Ridder et J. Theeuwes (éd.), Elsevier Science Publishers

Dibbs, R., Poulin, S. et Webber, M. (1994), «Utilisation des données fiscales dans l'Enquête sur la dynamique du travail et du revenu : rapport sommaire», Série des documents de recherche de l'EDTR, n° 94-11 au catalogue.

Fuller, W. (1987), *Measurement error models*, Wiley.

Groves, R. (1989), *Survey Errors and Survey Costs*, Wiley.

Michaud, S., Dolson, D. et Renaud, M. (1995), «Combinaison des données administratives et des données d'enquête en vue d'alléger le fardeau des répondants dans les enquêtes longitudinales», Série des documents de recherche de l'EDTR, n° 95-19 au catalogue.

Plewis, I. (1985), *Analysing change, Measurement and Explanation using longitudinal data*, Wiley.

Statistique Canada (1995), *Un aperçu de DAL, données administratives longitudinales. Rapports de DAL, Réf #94-20-01F, mai 1995. Préparé par la division des données régionales et administratives et la division des méthodes d'enquêtes sociales.* Statistique Canada.

L'ASSURANCE DE LA QUALITÉ DU REGISTRE CANADIEN DU CANCER

Leslie A. Gaudette, Tony LaBillois, Ru-Nie Gao et Heather Whittaker¹

RÉSUMÉ

Le Registre canadien du cancer (RCC) regroupe les données fournies par les 12 registres provinciaux ou territoriaux du cancer (RPTC) qui recueillent individuellement des données à partir de diverses sources administratives. La taille de la population observée, le nombre et la complexité des données élémentaires recueillies, ainsi que le raffinement des systèmes informatiques utilisés varient beaucoup d'un RPTC à l'autre. Le RCC est un fichier de données longitudinales axé sur la personne qui s'articule sur un système à trois modules, à savoir le module principal de contrôle, le module du couplage interne des enregistrements et le module de la confirmation des décès. Le module principal de contrôle n'accepte dans le RCC que les enregistrements qui satisfont un ensemble de contrôles rigoureux. Les enregistrements rejetés sont renvoyés aux RPTC concernés qui doivent les corriger, puis les transmettre de nouveau. Conjugué aux fonctions de mise à jour et de suppression, le module principal de contrôle assure que les données du RCC demeurent comparables à celles des bases de données dynamiques des RPTC. L'élaboration et la mise en oeuvre de processus d'assurance de la qualité est une composante conceptuelle importante du RCC. Cette composante comprend trois volets fondamentaux, à savoir la mise en place d'une infrastructure, y compris le développement de méthodes et d'outils, l'encouragement de la communication entre les parties intéressées et la mise en oeuvre par étape de stratégies de mesure du rendement. Ces trois volets ont été élaborés en s'inspirant des six attributs principaux de l'assurance de la qualité définis pour les registres du cancer. La création d'un Comité de la qualité des données (CQD), qui fait des recommandations sur la normalisation des données et encourage la communication régulière entre les parties intéressées, est une caractéristique essentielle du projet. La question de l'échantillonnage d'acceptation est résolue grâce à l'exécution, au niveau des RPTC, de contrôles rigoureux définis dans le Dictionnaire de données du RCC et dans l'ébauche du Manuel de procédures qui l'accompagne. En vue de contrôler les processus, on suit des lignes directrices sur la surveillance globale qui précisent, pour les données élémentaires clés, la fourchette acceptable quant à la proportion de valeurs inconnues ou de combinaisons improbables. À titre d'illustration, on présente les lignes directrices établies pour certains indicateurs de qualité et on montre combien il est important de les suivre pour répondre aux exigences des utilisateurs nationaux et internationaux.

MOTS CLÉS: Assurance de la qualité; registre du cancer; données administratives; erreurs non dues à l'échantillonnage.

1. INTRODUCTION

Le cancer est un problème de santé de plus en plus important aussi bien dans les pays développés que dans les pays en voie de développement. Au Canada, il est la deuxième cause la plus importante de décès, après les maladies cardiovasculaires. La probabilité globale (au cours de la vie) qu'un Canadien soit atteint du cancer est égale à un sur trois, et celle qu'il en meurt, à un sur quatre (Institut national du cancer du Canada, 1996). Le Canada est un des rares pays, et un des plus grands, où l'on a créé une base de données spéciale, appelée registre du cancer, afin de fournir les renseignements indispensables pour lutter efficacement contre le cancer (Statistique Canada, 1997; Band et coll., 1993). Les données des registres du cancer permettent de quantifier l'importance de la maladie, de décrire les caractéristiques de diverses formes de cancer et d'appuyer toute une gamme de projets de recherche, comme les travaux récents sur les origines génétiques du cancer, les études cas-témoins ou les études de cohortes, ainsi que les analyses économiques nécessaires à la planification des soins de santé.

1

Leslie A. Gaudette, analyste principale de recherche, Division des statistiques sur la santé, Statistique Canada, Ottawa, K1A 0T6, courrier électronique : lgaudet@statcan.ca. Tony LaBillois, méthodologiste principal, Division des méthodes d'enquêtes des ménages, Statistique Canada, Ottawa, K1A 0T6, courrier électronique : labiton@statcan.ca. Ru-Nie Gao, Unité du soutien à la recherche, Division des statistiques sur la santé, Statistique Canada, Ottawa, K1A 0T6. Heather Whittaker, directrice, Records and Registry, Manitoba Cancer Treatment and Research Foundation, Winnipeg (Manitoba), R3E 0V9, courrier électronique : heather@mctrf.mb.ca.

En 1969, le Canada a établi le Système national de déclaration des cas de cancer (SNDCC) pour recueillir, à l'échelle de la population, des données sur l'incidence du cancer. Vers le milieu des années 80, d'aucuns ont recommandé de créer un nouveau Registre canadien du cancer (RCC) afin de produire, comme le demandait un nombre croissant de chercheurs, des données sur le cancer plus précises, de meilleure qualité et d'une plus grande actualité. On a donc créé une base de données normalisées, axée sur la personne, pouvant être mise à jour, afin de fournir les données sur l'incidence du cancer et sur la survie à cette maladie nécessaires à la lutte anti-cancéreuse au Canada. On s'appuie, pour exploiter cette nouvelle base de données, sur un ensemble complexe de concepts et de méthodes d'assurance de la qualité qui permettent de répondre de sa fiabilité et de sa comparabilité à des bases de données semblables, tant au Canada qu'à l'étranger.

Dans le présent article, nous décrivons d'abord les forces à l'origine de la création du nouveau RCC, puis les principales composantes de ce dernier et l'infrastructure d'assurance de la qualité, y compris certains contrôles des processus mis au point récemment, et nous concluons par certaines réflexions générales et la description de travaux qui seront entrepris dans l'avenir.

2. CONTEXTE

2.1 Bref historique de la déclaration des cas de cancer au Canada

Les registres mis en place en Saskatchewan et en Colombie-Britannique autour de 1930 comptent parmi les plus anciens du monde. Nombre des autres provinces ont créé un registre au cours des années 40 et 50 (Band et coll., 1993). La plupart des registres provinciaux ou territoriaux du cancer (RPTC) ont été établis par l'organisme de lutte contre le cancer chargé de dispenser les soins anti-cancéreux dans la province ou le territoire et certains l'ont été par le ministère provincial ou territorial de la Santé. Dès le milieu des années 60, les progrès réalisés au Canada en matière d'enregistrement des cas de cancer ont rendu faisable la création d'un système national.

Par conséquent, en 1969, l'Institut national du cancer du Canada et le Bureau fédéral de la statistique (devenu aujourd'hui Statistique Canada), ont créé conjointement le Système national de déclaration des cas de cancer (SNDCC). Première base nationale sur l'incidence du cancer au Canada, le SNDCC regroupe les données collectées par les RPTC pour les cancers diagnostiqués de 1969 à 1991. Enfin, en 1989, on a entrepris la mise au point du RCC en vue d'améliorer la normalisation, la fiabilité et la qualité des données sur l'incidence du cancer (Statistique Canada, 1997).

2.2 Sources de données consultées par les registres provinciaux et territoriaux du cancer (RPTC)

Au Canada, les services de santé et la déclaration des cas de cancer relèvent des autorités provinciales ou territoriales. Dans la plupart des provinces et territoires, la loi exige que les cas de cancer soient déclarés à l'organisme approprié. Toutefois, la déclaration demeure volontaire à Terre-Neuve, au Nouveau-Brunswick et en Ontario. Chaque registre s'efforce d'enregistrer tous les nouveaux cas de cancer qui surviennent dans la population de sa province ou de son territoire en regroupant les renseignements en provenance de toutes les sources existantes. Ces dernières incluent les dossiers et les rapports de radiothérapie des centres anti-cancéreux, les dossiers médicaux des hôpitaux, des services de consultations externes et des cliniques privées, les rapports des laboratoires d'anatomo-pathologie et autres, ainsi que les rapports d'autopsie, les rapports des services de radiologie et des programmes de dépistage, les rapports des médecins exerçant en cabinet privé et les rapports sur la mortalité due au cancer des officiers de l'état civil (Statistique Canada, 1997; Band et coll., 1993).

2.3 Surdénombrement et sous-dénombrement

La comparabilité des données dépend à la fois du sous- et du surdénombrement, lesquels sont fonction de la source des données et des méthodes de traitement. On estime que, à l'heure actuelle, au niveau national, la couverture d'enregistrement atteint ou excède la norme acceptée de 95 % (Howe et coll., 1996), quoiqu'on ait observé une couverture plus faible dans certaines provinces, particulièrement durant les premières années (Band et coll., 1993). Il y a surdénombrement quand deux RPTC enregistrent le même cas sans le savoir ou que les identificateurs figurant dans les fichiers provinciaux ne suffisent pas au dépistage de toutes les déclarations en double. Le SNDCC

étant un système d'enregistrement de l'incidence du cancer, les cas n'y sont pas reliés à l'échelon national en vue de déceler les patients atteints de plusieurs tumeurs primitives. Avant 1992, les enregistrements en double dus à la déclaration d'un même patient par plus d'une province n'étaient pas systématiquement supprimés (sauf pour les deux territoires). Pour d'autres exemples de surdénombrement, consulter Statistique Canada (1997) ou Band et coll. (1993).

Le sous-dénombrement survient quand un registre n'utilise pas suffisamment de sources distinctes pour recenser les nouveaux cas de cancer. Certaines formes de cancer sont difficiles à diagnostiquer à cause de leur emplacement (ou siège) dans l'organisme. D'autres, tels que les cancers du poumon, du pancréas et de l'estomac, entraînent rapidement la mort. Un registre qui n'utilise pas les certificats de décès comme source de données finit généralement par s'apercevoir que ces formes de cancer sont sous-dénombrées. La déclaration incomplète des cas de cancer traités ailleurs que dans la province ou le territoire de résidence, l'utilisation de définitions différentes du néoplasme malin et la déclaration incomplète des enregistrements tardifs (ou des cas diagnostiqués après qu'un registre ait déclaré les cas de l'année) à Statistique Canada sont d'autres sources de sous-dénombrement (Statistique Canada, 1997).

2.4 Autres écarts entre RPTC ayant une incidence sur la qualité des données

La comparabilité des taux d'enregistrement des cas, d'incidence du cancer et de survie à la maladie, ainsi que la précision et la complétude des renseignements fournis varient selon les méthodes de codage, les définitions du cancer et les méthodes de saisie ou de traitement des données adoptées par les RPTC. Certains collectent plus de données élémentaires que d'autres et certains fournissent des données plus complètes et plus précises pour certaines variables liées au cancer. La rapidité de production des données est un autre problème important. En effet, si nombre de RPTC transmettent leurs données de six à 12 mois après la fin de l'année de référence, quelques-uns n'arrivent à les produire que trois à quatre ans plus tard. Certains RPTC sont équipés de systèmes informatiques très raffinés et très souples. Cependant, les grands registres doivent utiliser de gros systèmes fort complexes qui nécessitent parfois une reprogrammation importante, tandis que certains petits registres ne disposent que des systèmes informatiques de capacité insuffisante pour intégrer tous les contrôles automatisés qu'exige le nouveau système du RCC. Par conséquent, la qualité des données nationales dépend, en partie, de la capacité des systèmes informatiques des RPTC à satisfaire les nouvelles exigences de déclaration et de vérification. Néanmoins, grâce à la mise en oeuvre du nouveau RCC, les écarts entre RPTC s'amenuisent, particulièrement en ce qui concerne les éléments et les pratiques ayant une incidence sur les données nationales.

3. REGISTRE CANADIEN DU CANCER (RCC)

Le RCC est une base de données longitudinales normalisées, axée sur la personne, qui comprend des mécanismes permettant de mettre les enregistrements à jour, de déceler les enregistrements en double et (dans un avenir proche) d'effectuer la confirmation des décès à l'échelle nationale et qui regroupe les données sur les cas de cancer diagnostiqués au Canada depuis 1992. Le RCC est le point culminant de nombreuses années de collaboration entre les RPTC, Statistique Canada et nombres d'organisations qui se consacrent à la recherche sur le cancer au Canada.

3.1 Vue d'ensemble du Registre canadien du cancer

Le RCC comprend trois modules. Le module principal de contrôle exécute l'ensemble des procédures informatisées qui permettent de lire, de vérifier, de charger, de mettre à jour et de maintenir la base de données. Le module de couplage interne des enregistrements décele les enregistrements patient ou tumeur qui figurent éventuellement en double dans le RCC. Le module de confirmation des décès apparie ou couple les enregistrements patient du RCC aux enregistrements éventuellement correspondants de la Base canadienne de données sur la mortalité (BCDM) (base de données qui regroupe les renseignements tirés de toutes les déclarations de décès au Canada). Après l'exécution du module de confirmation des décès, les données du RCC peuvent être utilisées pour effectuer des études de survie et pour contrôler la taille de la base de données grâce au stockage séparé des données sur les cas dormants.

3.2 Préparation des données par les RPTC en vue de leur transmission au RCC

Avant de transmettre les données au RCC, les RPTC doivent les présenter sous une forme normalisée et enregistrer les données sur les nouveaux patients ou sur les nouvelles tumeurs en se servant des clichés d'enregistrement, des codes et des contrôles normalisés définis dans le Dictionnaire de données du RCC. Contrairement à celles du SNDCC, les données du RCC ne sont ni reformatées ni recodées par Statistique Canada. En lieu et place, le système du RCC exécute un ensemble de contrôles de validation et de cohérence rigoureux. Les enregistrements qui ne passent pas ces contrôles avec succès (habituellement, moins de 1 %) sont rejetés et renvoyés aux RPTC déclarants afin que ceux-ci les corrigent, puis les transmettent de nouveau à Statistique Canada. Seules les données satisfaisant tous les critères de vérification sont enregistrées dans la base de données du RCC. Les fichiers provinciaux ou territoriaux contenant un pourcentage trop élevé d'erreurs sont renvoyés au complet aux RPTC concernés afin que ceux-ci les corrigent, puis les transmettent de nouveau. Les données enregistrées dans le RCC doivent passer avec succès environ 80 contrôles de validation et de cohérence. Le RCC et les RPTC exécutent également quelques contrôles d'avertissement en vue de détecter les enregistrements qui contiennent des combinaisons de valeurs nécessitant un examen plus approfondi. Chaque année, tous les RPTC transmettent leur fichier d'enregistrements patient et tumeur à Statistique Canada au moment d'un cycle de traitement ordinaire ou d'un cycle de correction. Statistique Canada se sert alors du fichier électronique contenant les enregistrements pour entrer les données dans le RCC.

3.3 Module principal de contrôle

Ce module est le coeur du registre. Il comprend la base de données et toutes les fonctions qui y sont associées. Le RCC contient deux bases de données SAS couplées, l'une contenant les enregistrements patient et l'autre, les enregistrements tumeur. Comme le RCC est une base de données axée sur la personne, les données en entrée sur la personne atteinte du cancer figurent dans l'enregistrement patient et celles sur les caractéristiques du cancer, dans l'enregistrement tumeur. Chaque enregistrement patient contient un numéro d'identification unique du RCC et un numéro d'identification provincial qui permettent de le coupler aux enregistrements tumeur pertinents enregistrés dans la base de données.

Les types d'enregistrements patient et tumeur que les RPTC peuvent transmettre à Statistique Canada aux fins de l'entrée dans le module principal de contrôle du RCC sont les suivants :

- *Nouvel enregistrement* : Un RPTC présente un nouvel enregistrement patient quand le patient ne figure pas encore dans la base de données. Cet enregistrement contient tous les renseignements personnels concernant le patient. Un nouvel enregistrement tumeur est transmis pour toute nouvelle tumeur diagnostiquée chez un patient donné.
- *Enregistrement de mise à jour* : Un enregistrement complet de mise à jour doit être envoyé chaque fois qu'une RPTC veut modifier une ou plusieurs zones d'un enregistrement patient ou tumeur déjà enregistré dans le RCC.
- *Enregistrement de suppression* : Un RPTC peut supprimer un enregistrement patient ou tumeur en transmettant un enregistrement de suppression contenant certaines zones exactement identiques à celles qui apparaissent dans l'enregistrement déjà contenu dans la base de données du RCC. Pour supprimer un enregistrement tumeur, l'envoi d'un seul enregistrement de suppression suffit, mais pour supprimer un enregistrement patient, le RPTC doit transmettre des enregistrements de suppression correspondant à l'enregistrement patient ainsi qu'à tous les enregistrements tumeur pertinents.
- *Enregistrement de changement de propriété s'appliquant à un enregistrement patient* : Quand une nouvelle tumeur est diagnostiquée chez un patient déjà enregistré et que le diagnostic est posé dans une autre province que celle où a été diagnostiquée la tumeur précédente, le RPTC qui effectue l'enregistrement de la nouvelle tumeur peut présenter un enregistrement de changement de propriété pour devenir le propriétaire de l'enregistrement patient.

Les règles de contrôle intégrées au module sont définies dans le Dictionnaire de données du RCC et représentent la première étape de l'assurance de la qualité des données. Les routines de validation sont appliquées pour vérifier le contenu de certaines zones des enregistrements. Les contrôles de cohérence permettent de vérifier la cohérence des données élémentaires qui figurent dans l'ensemble d'enregistrements patient et tumeur soumis pour chaque patient. Les contrôles d'appariement des enregistrements d'entrée permettent d'assurer que pour chaque envoi, l'ensemble des enregistrements d'entrée pour un patient est complet et forme un tout cohérent par rapport à

l'opération à effectuer (par exemple, inscription de nouveaux enregistrements, modifications ou suppressions d'enregistrements existant dans le RCC et (ou) changement de propriété d'un enregistrement patient). Enfin, des règles de contrôle supplémentaires sont appliquées pour s'assurer que les changements apportés au RCC soient conformes à la structure et à la logique interne du registre.

Le module principal de contrôle présente un grand nombre de nouvelles caractéristiques conceptuelles. Ainsi, le RCC est basé sur la deuxième édition de la Classification internationale des maladies-Oncologie (CIM-O). Les tumeurs primaires multiples sont définies d'après un ensemble de règles visant les codes à 4 chiffres de la CIM-O-Topographie (siège du cancer), les groupes de codes de la CIM-O-Morphologie (type histologique) et les codes de latéralité. Des programmes assurant la conversion automatique des codes d'une version de la CIM à ceux d'une autre sont en place à l'échelon national (RCC). Les enregistrements marqués d'un signal d'avertissement à la suite de l'exécution de ces programmes sont renvoyés aux RPTC déclarants afin que ceux-ci les examinent et soumettent tout enregistrement de mise à jour nécessaire. En outre, les RPTC disposent chacun d'un logiciel de conversion des codes de la CIM-O et des codes postaux qui leur permet d'effectuer automatiquement la conversion aux codes de la Classification géographique type jusqu'aux niveaux de la région métropolitaine de recensement et du secteur de recensement.

Deux semaines après avoir reçu les données des RPTC, Statistique Canada envoie à tous les registres des rapports de rétroaction contenant des renseignements sommaires, ainsi que des renseignements détaillés (par enregistrement) sur les enregistrements rejetés lors de certains contrôles. Des messages d'avertissement ne sont produits que pour quelques contrôles seulement. Un ensemble de rapports de contrôle de la qualité permet de surveiller la complétude de la déclaration de chaque donnée élémentaire, ainsi que les indicateurs de la qualité des données. Chaque RPTC est tenu d'examiner ces rapports, de corriger ses enregistrements au besoin et de présenter tout enregistrement nouveau ou de mise à jour résultant au moment de l'envoi suivant d'enregistrements au RCC.

Le dépistage des enregistrements en double s'effectue de la façon suivante. Chaque RPTC doit transmettre tout enregistrement concernant un résident d'une autre province ou d'un autre territoire au RPTC pertinent, car le système du RCC n'accepte que les enregistrements tumeur pour lesquels le RPTC déclarant est celui de la province ou du territoire de résidence. À cet égard, le module des contrôles de base vérifie tous les enregistrements patient ou tumeur en entrée en se fondant sur les numéros d'identification provinciaux en vue de déceler les enregistrements en double. Puis, le module de couplage interne des enregistrements effectue un dernier dépistage des enregistrements patient ou tumeur qui figurent en double dans le RCC.

3.4 Module de couplage interne des enregistrements

En plus de réduire le surdénombrement, le module de couplage interne des enregistrements fournit aux RPTC des renseignements rétroactifs sur des tumeurs enregistrées par d'autres registres pour certains de leurs patients. Un enregistrement en double peut survenir quand une tumeur ou un patient est enregistré dans plus d'une des sources de renseignements consultées par un RPTC et que ce dernier attribue des numéros d'identification du patient distincts aux enregistrements. Un enregistrement en double peut aussi avoir lieu si une tumeur ou un patient est enregistré dans plus d'une province et que les RPTC concernés envoient chacun l'enregistrement au RCC, parce qu'il n'y a pas eu d'échange de renseignements au sujet du cas.

Chaque année, on arrête le cycle ordinaire de traitement des données du RCC pendant deux mois afin d'exécuter le couplage interne des enregistrements. Après avoir prétraité la base de données du RCC, le module effectue un couplage probabiliste des enregistrements destiné à comparer la base de données à elle-même. Le couplage probabiliste s'appuie sur des règles complexes qui permettent de mesurer le degré de similarité de variables clés qui, en soi, ne sont pas des identificateurs uniques, et d'attribuer des coefficients de pondération à ces variables. Si la somme des coefficients de pondération calculés pour une combinaison donnée d'enregistrements dépasse un seuil préétabli, Statistique Canada communique le groupe d'enregistrements en question aux RPTC concernés pour qu'ils l'examinent et déterminent s'il contient des enregistrements en double. La détection de groupes d'enregistrements patient ou tumeur éventuellement en double dans la base de données du RCC nécessite l'exécution d'un ensemble de programmes complexes. Pour chaque groupe repéré, le système produit un rapport de rétroaction qui est envoyé aux RPTC pertinents afin qu'ils proposent une solution. Environ quatre semaines sont accordées aux RPTC pour examiner les rapports et pour décider d'une solution, exercice qui peut nécessiter la consultation d'autres RPTC. Puis, les RPTC renvoient les rapports à Statistique Canada après y avoir indiqué

si, selon leur examen, les enregistrements patient ou tumeur visés sont ou non des enregistrements en double. La base de données du RCC est alors mise à jour en conséquence et des rapports de confirmation sont envoyés aux RPTC.

3.5 Situation actuelle du RCC

Le RCC est opérationnel depuis la fin de 1994. En décembre 1996, il contenait les données de tous les RPTC pour les années de référence 1992 et 1993, les données de la plupart pour 1994 et de deux pour 1995. Environ 120 000 nouveaux enregistrements patient et 125 000 nouveaux enregistrements tumeur sont reçus chaque année. Le premier cycle de couplage interne des enregistrements, exécuté en janvier et février 1996, a indiqué que la base de données contenait 0,2 % d'enregistrements en double, environ la moitié des doubles comptes ayant eu lieu entre RPTC distincts et le reste au sein d'un RPTC. Ce résultat montre que les RPTC dépistent avec succès les enregistrements en double avant de transmettre leurs données au RCC. Cependant, le nombre d'enregistrements en double survenant entre provinces pourrait augmenter car la probabilité que les patients déménagent d'une province à l'autre croît au fil du temps. Enfin, le développement et les tests du module de confirmation des décès sont maintenant bien en train et le module sera, en principe, exécuté pour la première fois en 1997.

4. ASSURANCE DE LA QUALITÉ

4.1 Assurance de la qualité, Registre canadien du cancer et Comité de la qualité des données

L'assurance de la qualité (AQ) est un élément conceptuel important du RCC qui permet de répondre aux exigences croissantes des utilisateurs quant à l'actualité et à la qualité des données. Les utilisateurs forment, en effet, un groupe de plus en plus diversifié allant des généticiens aux planificateurs des services de santé en passant par les épidémiologistes. Le recours aux méthodes d'assurance de la qualité vise aussi à optimiser les coûts en transférant la responsabilité des contrôles de validation aux RPTC, mesure qui diminue le coût de la production de routine à Statistique Canada et réduit au minimum le fardeau de réponse imposé aux fournisseurs de données. L'automatisation de certaines fonctions de codage diminue les coûts assumés tant par les RPTC que par Statistique Canada et procure du temps libre pour l'exécution d'autres activités, comme le couplage interne des enregistrements et la confirmation des décès. De surcroît, cette stratégie systématique d'assurance de la qualité réduit les coûts occultes inhérents à l'épuration fastidieuse des données de mauvaise qualité.

Le Comité de la qualité des données (CQD), comité permanent qui relève du Conseil canadien des registres du cancer (ou Conseil) est responsable de la gestion de la qualité des données du RCC. Le Conseil, qui se réunit au moins une fois par an depuis dix ans, est co-présidé par un représentant de Statistique Canada et compte parmi ses membres des représentants des RPTC et des groupes principaux d'utilisateurs. Le CQD fait des recommandations et donne son avis au Conseil sur des questions concernant la qualité et la normalisation des données collectées, codées et publiées par le RCC. Afin d'établir des positions par consensus, le Comité effectue périodiquement auprès des RPTC des sondages sur les méthodes et les pratiques courantes d'enregistrement dans les diverses régions du Canada. La mise en place d'une infrastructure englobant une série d'*outils*, de stratégies de *communication* et de méthodes de *mesure du rendement* permet de résoudre les problèmes de qualité selon des processus qui visent à améliorer divers *attributs* ou caractéristiques des données.

4.2 Outils

Le Dictionnaire de données du RCC définit les clichés d'enregistrement, les données élémentaires, les contrôles de validation, les contrôles de cohérence et les fichiers de référence et de conversion normalisés que doivent utiliser les RPTC pour soumettre leurs données au RCC. La normalisation poussée des données en entrée réduit les coûts d'exploitation grâce à la détection plus précoce des erreurs. Après chaque envoi de données au RCC, les RPTC reçoivent des rapports de rétroaction détaillés où sont décrits les problèmes relevés. Compte tenu de l'évolution de la prestation des soins médicaux, il convient de faire preuve de vigilance afin d'être certain que les cas diagnostiqués par de nouveaux fournisseurs de services médicaux, comme les services de consultations externes, et par les programmes de dépistage soient enregistrés. L'ébauche du Manuel des procédures du RCC contient les définitions de cas et la description des procédures d'enregistrement qu'il faut appliquer pour s'assurer que la

confirmation des cas atteint ou dépasse l'objectif de 95 %. Le CQD sert aussi d'agent de liaison avec la *North American Association of Central Cancer Registries (NAACCR)* et le Centre international de recherche sur le cancer (CIRC) avec lesquels il coordonne l'élaboration et l'application de normes pour les registres du Canada, des États-Unis et d'autres pays (*American Association of Central Cancer Registries*, 1994; Parkin et coll., 1992; Parkin et coll., 1994).

4.3 Communications

Une stratégie de communication énergique fait pendant aux méthodes techniques susmentionnées. Les employés des registres peuvent communiquer avec un spécialiste de la classification médicale ou avec d'autres membres du personnel du registre du cancer de Statistique Canada par téléphone, par télécopieur ou par courrier électronique, pour discuter de tout problème posé par le codage ou par la déclaration des données. Le bulletin intitulé *La filière du cancer*, qui est publié deux ou trois fois par an, est le principal instrument utilisé pour communiquer aux membres du personnel technique des registres du cancer et des hôpitaux les décisions concernant le codage et la déclaration des données. Ce bulletin facilite les échanges entre les codeurs de données, les membres du personnel technique de registres, les programmeurs et les médecins concernant les problèmes de qualité des données. Des ateliers techniques, organisés par les codeurs et les gestionnaires des données sur le cancer et destinés à ces derniers, sont tenus tous les 18 mois environ pour faire connaître les nouvelles procédures du RCC et pour examiner les questions ayant trait à l'amélioration de la qualité et au codage des données. De surcroît, des représentants de Statistique Canada rendent visite aux divers registres pour assurer la formation dans des domaines particuliers, tels que le codage en fonction des diverses classifications ou l'examen des procédures, de façon à éliminer les obstacles à la déclaration des données au RCC. Enfin, les membres tant du CQD que du Conseil font part de leurs recommandations et de leurs décisions dans les comptes rendus des conférences téléphoniques et des réunions annuelles. La mise en oeuvre de ces stratégies de communication a abouti à la création d'un réseau dynamique qui s'étend à tout le pays et facilite la détection et la résolution des problèmes.

4.4 Mesures du rendement

La stratégie d'assurance de la qualité appliquée au RCC se fonde sur les méthodes statistiques de contrôle de la qualité, à savoir l'échantillonnage pour acceptation, le contrôle des processus et les études spécialement conçues (Hilsenbeck, 1994). L'*échantillonnage pour acceptation* comprend l'inspection, et l'acceptation ou le rejet, des données brutes au moyen du Dictionnaire de données et du module principal de contrôle du système du RCC. Le *contrôle des processus* vise à surveiller la production, afin de déterminer quand des mesures correctives sont nécessaires (voir aussi la section 5.2). Les *études spécialement conçues*, qui représentent la méthode de contrôle de la qualité la plus élaborée (et la plus coûteuse), servent à évaluer la complétude de la confirmation des cas (exercice comparable à la contre-vérification des dossiers dans le cas du Recensement) ou à effectuer de nouveau l'abstraction des données de cas déjà enregistrées afin d'évaluer la précision et la complétude des données élémentaires déclarées.

4.5 Attributs des données

On peut évaluer la qualité des données des registres du cancer d'après certains attributs qui, en fait, représentent les principales sources d'erreurs non dues à l'échantillonnage. Ces attributs sont la confirmation des cas, la précision des données, la complétude des données, l'actualité des données et la constance (Hilsenbeck, 1994). La *confirmation des cas* a trait au sous-dénombrement aussi bien qu'au surdénombrement des cas (voir la section 2.3). D'après des mesures indirectes, les registres canadiens obtiennent de bons résultats dans ce domaine. Néanmoins, il est nécessaire d'effectuer des études spécialement conçues de la confirmation des cas afin d'obtenir des renseignements complets sur les domaines problématiques. La *précision* des données élémentaires enregistrées dans le RCC est contrôlée partiellement grâce aux contrôles rigoureux définis dans le Dictionnaire de données du RCC, ainsi qu'à une surveillance globale au moyen de tableaux de contrôle de la qualité. Toutefois, il faut également effectuer des études spécialement conçues de recodage et de réabstraction. Les tableaux de contrôle de la qualité permettent, en outre, d'évaluer la *complétude des données*, grâce à l'établissement de lignes directrices précisant les proportions acceptables de valeurs inconnues (voir aussi la section 5). L'*actualité des données* a trait au délai requis pour publier les données, délai qui a une incidence sur leur utilité, c'est-à-dire un attribut de qualité. On se rapproche maintenant des 20 mois fixés comme objectif par le Conseil canadien des registres du cancer pour

la production des données, en partie parce que le système du RCC permet de réduire considérablement le temps de traitement à l'interne. La *constance* des classifications utilisées pour le codage et des méthodes de publication est assurée grâce à une entente sur des dates de mise en oeuvre communes établies par le Conseil.

5. CONTRÔLE DES PROCESSUS DU RCC

5.1 Vue d'ensemble du contrôle des processus du RCC

Le CQD a élaboré des rapports et des lignes directrices sur le contrôle de la qualité qui permettent de fournir aux RPTC des renseignements rétroactifs grâce à la surveillance de la proportion de valeurs inconnues dans les zones clés et d'enregistrements douteux repérés en appliquant des algorithmes représentant divers indicateurs de qualité. Les lignes directrices ont été élaborées en collaboration avec plusieurs sources, après avoir examiné minutieusement les données d'entrée. Les totalisations de données fondées sur ces lignes directrices sont examinées par les employés de Statistique Canada en collaboration avec le CQD, en vue de dresser la liste des priorités quant aux corrections à apporter, donc, de contribuer à l'amélioration du système d'enregistrement des cas de cancer au Canada. Les RPTC sont tenus d'examiner leurs enregistrements en regard de certains indicateurs quand la proportion de cas douteux tombe en dehors des fourchettes recommandées. Les combinaisons d'enregistrements qui portent un signal signifiant que les RPTC ont déjà examiné ces cas ne nécessitent plus aucun examen. Les RPTC doivent également donner des précisions quand les résultats des contrôles se situent en dehors de la fourchette idéale ou acceptable, et apporter les corrections nécessaires, au besoin.

5.2 Exemples d'indicateurs

1. *Codes de siège/topographie de tumeur non spécifiques* : pourcentage d'enregistrements transmis au RCC pour lesquels le siège précis de la tumeur (p. ex., sein ou poumon) est soit inconnu, soit non précisé. Comme tous les RPTC respectent la norme proposée de moins de 5 %, aucune action supplémentaire n'est nécessaire (figure 1).

2. *Codes de morphologie (type histologique) de tumeur non spécifiques* : pourcentage d'enregistrements présentant un code de morphologie non spécifique transmis au RCC (p. ex., alors que le sarcome est une tumeur dont l'histologie est normalement typique des os, quelques cas rares se manifestent à d'autres sièges, comme le sein). Comme les résultats présentés à la figure 2 représentent la moyenne pour tous les sièges de cancer confondus, la proportion d'enregistrements pour lesquels la morphologie est inconnue pourrait être plus forte ou plus faible pour certains sièges de tumeur. Une proportion trop forte pourrait être le signe d'un degré inacceptable d'imprécision en ce qui concerne la déclaration ou le codage des renseignements diagnostics, tandis qu'une proportion faible pourrait indiquer que le registre concerné n'utilise pas les sources peu fiables de données. Les registres pour lesquels la valeur de l'indicateur tombe au-delà de la fourchette acceptable de 0 à 8 % s'efforcent d'améliorer la complétude des renseignements qui figurent dans les enregistrements sources (figure 2).

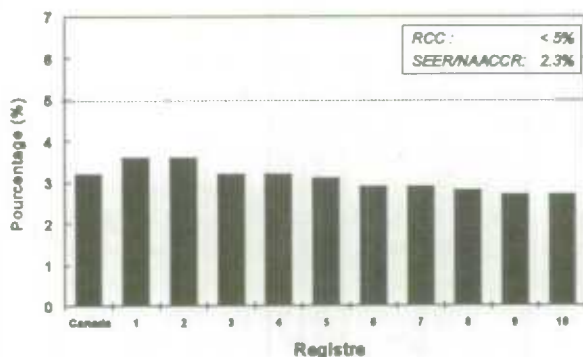


Figure 1. Pourcentage de codes de siège non spécifiques

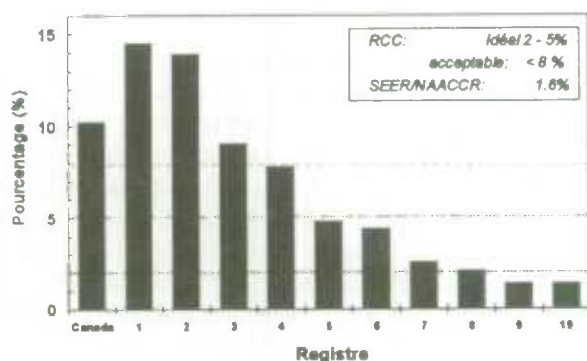


Figure 2. Pourcentage de codes de morphologie non spécifiques

Dans les figures 1 à 6, la première barre du graphique représente le taux national et les dix autres barres représentent les dix registres provinciaux classés par ordre décroissant pour chaque indicateur. Comme le rang varie d'un indicateur à l'autre, il est impossible de comparer les registres. Les lignes en trait interrompu représentent le niveau idéal recommandé, tandis que les lignes en pointillé qui figurent sur certains graphiques indiquent le niveau ou la fourchette acceptable. L'analyse est fondée sur les données de 1992 du Registre canadien du cancer.

3. Combinaisons siège-morphologie-âge improbables : Cet indicateur est un exemple de plusieurs contrôles des processus du RCC destinés à déceler les combinaisons de valeurs improbables ou inhabituelles. Par exemple, le cancer de la prostate ne se manifesterait que rarement, voire jamais, chez un garçon de cinq ans - l'âge réel est probablement 105 ans! Chaque registre est tenu d'examiner tous les enregistrements tumeur pour lesquels la combinaison siège-morphologie-âge est improbable. Les RPTC qui ne satisfont pas la norme de moins de 0,1 % doivent passer leurs méthodes en revue, déterminer les causes du pourcentage trop élevé et prendre les mesures correctives nécessaires. Certains registres vérifient ces combinaisons en direct et marquent les enregistrements visés d'un indicateur signifiant que les combinaisons inhabituelles ont été vérifiées. Dans l'ensemble, environ cinq combinaisons particulières représentent les deux tiers des combinaisons improbables (Gao et Gaudette, 1994) (figure 3).

4. Proportion de certificats de décès uniquement (%CDU) : Proportion d'enregistrements pour lesquels la seule source de renseignements diagnostiques est le certificat de décès. Ce contrôle est un moyen indirect important d'évaluer la complétude des travaux de confirmation des cas. Les résultats semblent s'uniformiser d'un registre à l'autre au fil du temps et nombre de registres atteignent maintenant, ou sont près d'atteindre, un taux de CDU compris dans la fourchette idéale. Un des registres n'effectue pas encore la confirmation des décès. En revanche, le taux de CDU d'un autre est passé de plus de 10 % à moins de 5 % au cours des dix dernières années (figure 4).

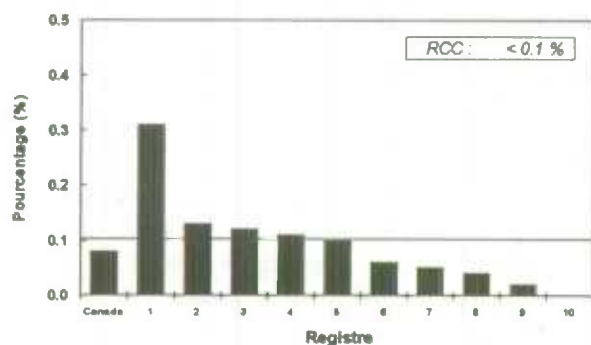


Figure 3. Pourcentage de combinaisons siège-morphologie-âge improbables

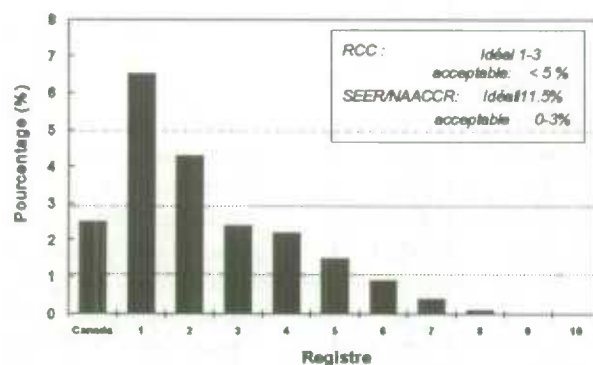


Figure 4. Pourcentage de certificats de décès uniquement

5. Proportion de divisions de recensement inconnues : En ce qui concerne les données géographiques, la ligne directrice pour les données manquantes a été établie à moins de 5 % compte tenu des besoins analytiques. Les RPTC ont accès à des systèmes informatisés de codage géographique qui réduisent les coûts et améliorent la complétude de la déclaration des données géographiques. Les registres pour lesquels on observe une forte proportion de valeurs manquantes doivent s'efforcer de consulter d'autres fichiers provinciaux contenant les codes postaux et d'autres renseignements sur les lieux de résidence afin d'améliorer la complétude des données pour l'élément de donnée visé (figure 5).

6. Proportion de données inconnues (p. ex., jour de la naissance) : La ligne directrice pour le jour de la naissance, fixée au départ à moins de 1 %, a été augmentée à moins de 5 %, norme que tous les registres, sauf deux, satisfont. La norme a été modifiée parce que certains RPTC imputent systématiquement le 1^{er} janvier, le 30 juin ou le 1^{er} juillet d'une année donnée pour les dates inconnues, ou le 15 d'un mois connu pour les jours inconnus, c'est-à-dire des dates qui sont toutes valides dans le RCC. Les deux registres pour lesquels la proportion est supérieure

à 5 % marquent d'un repère les dates imputées et convertissent les codes à la valeur correcte pour les dates inconnues au moment de transmettre les enregistrements au RCC. Donc, un faible pourcentage de jours de naissance inconnus ne reflète pas nécessairement le pourcentage réel de valeurs inconnues dans un registre donné, si ce pourcentage est fondé sur des données imputées, tandis qu'un pourcentage élevé témoigne parfois d'une méthode de déclaration plus précise (figure 6).

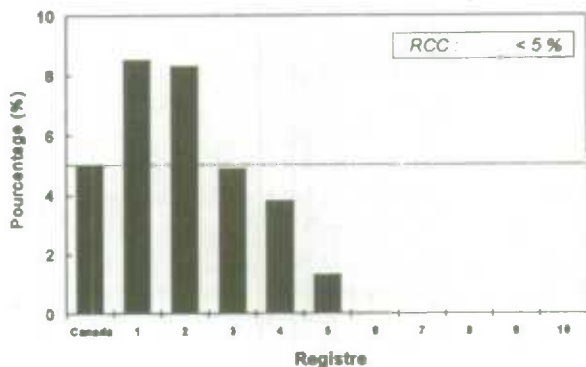


Figure 5 Division de recensement : pourcentage de valeurs inconnues

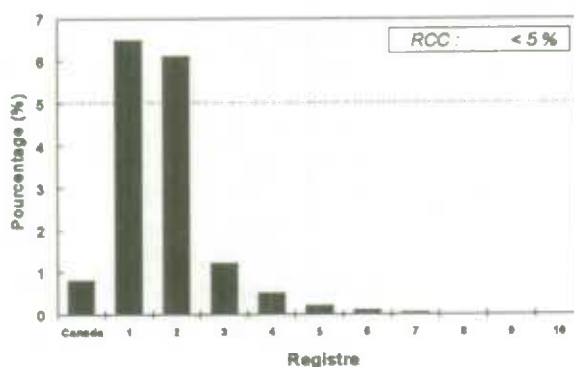


Figure 6. Date de naissance (jour) : pourcentage de valeurs inconnues

6. RÉSUMÉ ET ORIENTATIONS FUTURES

Afin de satisfaire les besoins divergents des utilisateurs, il est indispensable de surveiller en permanence la qualité des données et de mettre en oeuvre de nouvelles mesures destinées à résoudre les problèmes de qualité. Bien que la vitesse de production des données augmente, les retards de déclaration demeurent une source de préoccupation. Les chercheurs canadiens et étrangers ont besoin de données de haute qualité pour justifier les demandes de bourse et pour fournir les renseignements permettant d'établir la validité des résultats de leurs travaux. Parallèlement, il devient de plus en plus important de produire des données mieux uniformisées à l'échelle internationale, de sorte que leur interprétation soit significative et que l'on puisse adopter une stratégie internationale de recherche et d'élaboration des programmes dans le domaine de la lutte anticancéreuse. En outre, il est indispensable de gérer prudemment les ressources de moins en moins importantes allouées pour la mise au point et l'exploitation des registres du cancer et de définir les priorités afin d'optimiser les coûts du contrôle de la qualité et de la production.

À l'heure actuelle, l'échantillonnage d'acceptation en vue d'assurer la qualité des données en entrée est exécuté de manière assez efficace grâce à plusieurs outils, dont le Dictionnaire de données du RCC. Les rapports de rétroaction et de contrôle de la qualité produits par le système du RCC permettent de dépister efficacement les problèmes de vérification ainsi que les problèmes de qualité plus généraux. Enfin, les lignes directrices sur le contrôle des processus ont été approuvées par le Conseil en mai 1996. Cette stratégie globale d'assurance de la qualité, qui a été mise en oeuvre après consultation approfondie des employés des registres provinciaux, territoriaux et nationaux appartenant à diverses disciplines et avec la participation de ces derniers, représente un grand pas vers l'avant en ce qui a trait à la qualité des données produites par les registres du cancer au Canada.

La confirmation des cas demeure problématique. À l'heure actuelle, le CQD élabore, en collaboration avec la NAACCR et des groupes d'utilisateurs clés, un plan d'étude spécialement conçue de la complétude de l'enregistrement des données sur l'incidence du cancer qui s'inspire de l'étude mise au point par la NAACCR. Par ailleurs, le RCC ne contient pas de données sur la détermination du stade d'évolution des tumeurs ni sur le traitement, données qui sont de plus en plus nécessaires pour surveiller les efforts de lutte anticancéreuse et évaluer les résultats des traitements. Enfin, le RCC et, en fait, tous les registres du cancer doivent s'adapter de plus en plus

rapidement à l'évolution de l'environnement, notamment aux changements d'ordre politique, médical, technologique et économique.

7. BIBLIOGRAPHIE

American Association of Central Cancer Registries. (1994). *Standards for Cancer Registries Volume III. Standards for Completeness, Quality, Analysis and Management of Data*. Seiffert, J.E. (ed.) Sacramento, CA: NAACCR Cancer Surveillance and Control Program.

Band, P.R., Gaudette, L.A., Hill, G.B., Holowaty, E.J., Huchcroft, S.A., Makomaski Illing, E.M., Mao, Y., et Semenciw, R.M. (1993). *Développement du registre canadien du cancer: Incidence du cancer au Canada et dans les 6 régions canadiennes, 1969 à 1988*. Ottawa: Conseil canadien des registres du cancer, Santé et Bien-être social Canada, Statistique Canada.

Gao, R.N., et Gaudette, L.A. (1994). *Assessment of edits for improbable or impossible combinations of site, morphology and age for Canadian cancer registries*. Ottawa: Division des statistiques de la santé, Statistique Canada.

Hilsenbeck, S.G. (1994). Quality Control. In: Menck H., and Smart C., (eds.), *Central Cancer Registries: Design, Management and Use*, pp 131-177. USA: Harwood Academic Publishers.

Howe, H.L., Lehnerr, M., Derrick, L., et coll. (1996) *Cancer Incidence in North America 1988-1992*. Sacramento, CA: North American Association of Central Cancer Registries.

Institut national du cancer du Canada (1996). *Statistiques canadiennes sur le cancer 1996*. Toronto: Institut national du cancer du Canada.

Parkin, D.M., Chen, V.W., Ferlay, J., Galceran, J., Storm, H.H., et Whelan, S.L. (1994). *Comparability and Quality Control in Cancer Registration*. Lyon: International Agency for Research on Cancer. IARC Technical Report No. 19.

Parkin, D.M., Muir, C.S., Whelan, S.L., Gao, Y.T., Ferlay, J., et Powell, J., (eds.), (1992). *Cancer Incidence in Five Continents, Volume VI*. Lyon: International Agency for Research on Cancer, IARC Scientific Publications No. 120.

Statistique Canada (1997). *L'incidence du cancer au Canada 1969-1993*. Ottawa: Statistique Canada, catalogue 82-566 (hors série).

INCIDENCE DE LA RÉGLEMENTATION ET DES PRATIQUES ADMINISTRATIVES SUR LE FONCTIONNEMENT D'UN SYSTÈME STATISTIQUE AXÉ SUR UN REGISTRE

- CAS DU REGISTRE DES PRESTATAIRES DE L'ASSURANCE-CHÔMAGE

Timo Koskimäki¹

RÉSUMÉ

Le moyen le plus courant d'évaluer les statistiques fondées sur des dossiers administratifs consiste à les comparer aux macrodonnées d'une enquête ou d'un recensement. Dans la présente étude, on évalue aussi la qualité des statistiques administratives en les couplant aux microdonnées d'une enquête à grande échelle et en déterminant si les données du registre concerné sont cohérentes au fil du temps. Bien que la comparaison au niveau des macrodonnées donne à penser que les données administratives sont acceptables, le contrôle de cohérence et le couplage à des données d'enquête indiquent que la qualité du registre étudié varie considérablement. Comme les erreurs s'annulent dans une certaine mesure, le registre des prestataires de l'assurance-chômage est satisfaisant en tant que macro-indicateur du chômage. Toutefois, la variabilité de la qualité des données peut poser des problèmes quand on effectue une analyse de microdonnées ou qu'on se sert des données du registre comme données auxiliaires pour calculer des estimations à partir de données d'enquête.

MOTS CLÉS: données administratives; qualité; registre; enquête sur la population active.

1. CONTEXTE

L'évaluation empirique de la qualité des statistiques administratives est souvent difficile à réaliser. On peut appliquer trois méthodes fondamentalement différentes. La plus courante consiste à comparer les statistiques administratives à des macrodonnées auxiliaires choisies comme référence. Par ailleurs, si on dispose de données empiriques et qu'on peut résoudre les problèmes de protection des données, on peut effectuer une comparaison de microdonnées. Enfin, on peut étudier les pratiques administratives qui sous-tendent un registre et essayer de mesurer leur effet sur les statistiques établies d'après les données de ce registre.

Le présent article a pour objet d'étudier, sous les trois angles susmentionnés, les facteurs qui ont une incidence sur le contenu et sur la qualité du registre (finlandais) des personnes à la recherche d'un emploi, c'est-à-dire le dénombrement des prestataires de l'assurance-chômage. L'enquête mensuelle sur la population active sert d'enquête de référence et de source de microdonnées.

Il convient avant tout de comparer les mesures en tant qu'indicateurs économiques à court terme. D'autres aspects du chômage - par exemple, les problèmes sociaux et individuels qu'il pose - tombent hors du champ de la présente étude.

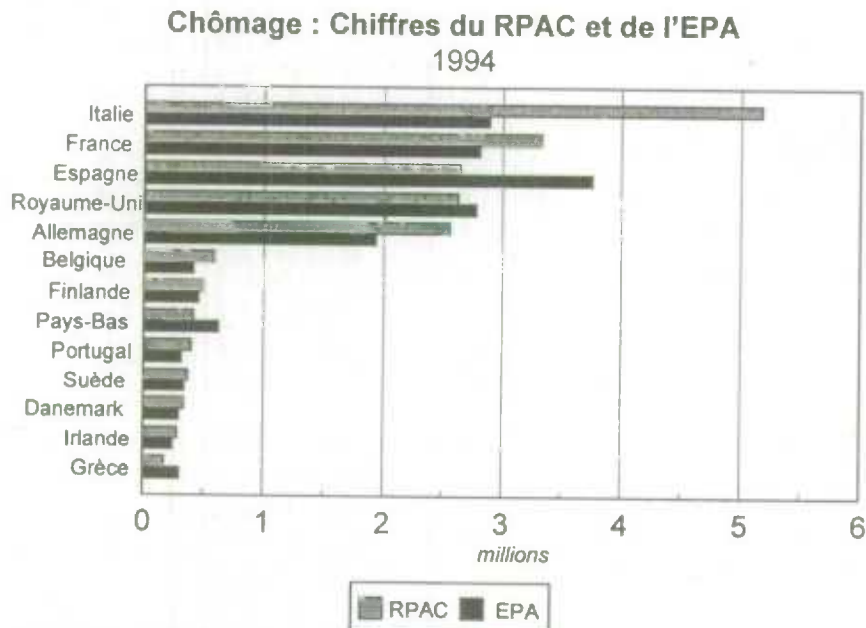
¹ Timo Koskimäki, Statistics Finland, P.O.B. 5B FIN-00022 Statistics Finland
E-mail: timo.koskimaki@stat.fi

2. LES DEUX MESURES DU CHÔMAGE

Dans la plupart des pays, il est courant de produire deux mesures officielles du chômage, l'une basée sur une source de données administratives comme le Registre des prestataires de l'assurance-chômage (RPAC) et l'autre, sur des données d'enquête comme l'enquête sur la population active (EPA). Les deux mesures pourraient indiquer des niveaux de chômage fort semblables, mais, très souvent, il en est autrement. La figure 1 montre que la concordance des estimations du chômage basées sur les données du RPAC et sur celles de l'EPA est discutable. Dans certains cas, le niveau de chômage indiqué par l'EPA est considérablement plus faible que celui indiqué par le registre. Dans d'autres, c'est l'inverse qu'on observe.

La raison principale de ces écarts est assez évidente : la couverture, le contenu et la qualité des statistiques basées sur le registre dépendent de l'infrastructure du système statistique, c'est-à-dire la réglementation, ainsi que les administrations centrale et locales. Fondamentalement, on peut donc voir dans la normalisation des enquêtes sur la population active un effort en vue de résoudre le problème que pose la comparaison des statistiques sur le chômage entre régions et, ultérieurement, entre pays (Innes 1990).

Figure 1 :



Source : Eurostat: Unemployment 2/95; Statistics Finland; Statistics Sweden

2.1 Cas de la Finlande

En Finlande, les deux mesures du chômage sont, à première vue, étonnamment proches (voir l'annexe 1). Toutefois, en examinant les séries chronologiques, on s'aperçoit qu'au-delà de la concordance générale, la correspondance entre les deux mesures varie au cours du temps. La concordance est surtout forte durant la période de dix ans allant du milieu des années 70 au milieu des années 80. Avant cela, l'enquête sur la population active indiquait un niveau de chômage nettement plus élevé que le registre. En revanche, depuis le milieu des années 80, c'est l'inverse que l'on observe (figure 2).

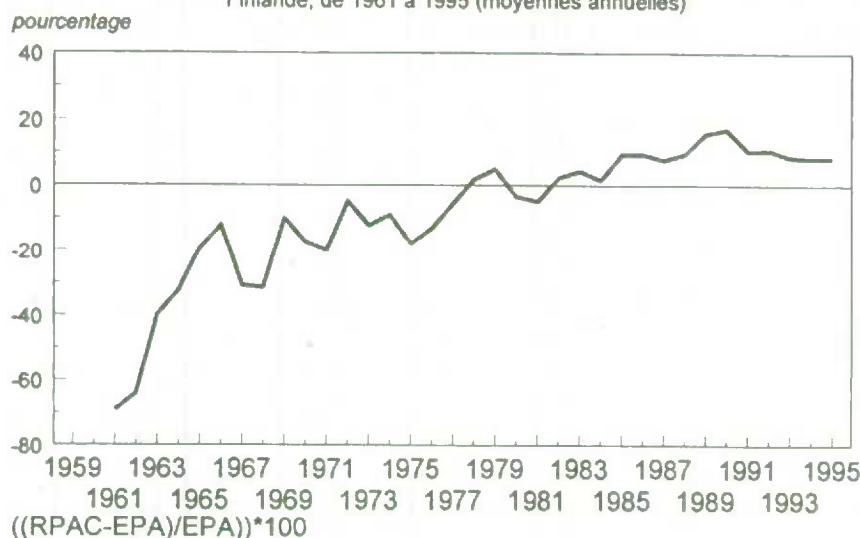
La concordance des deux séries de données pendant la décennie susmentionnée tient à ce que, durant cette période, la couverture du régime d'assurance-chômage et les ressources de l'administration publique étaient suffisantes. On avait achevé de mettre en place le régime général d'assurance-chômage et les autorités responsables de l'emploi accordaient des prestations d'assurance-chômage fort généreuses. Parallèlement, le système administratif est devenu plus centralisé et plus efficace, en ce qui concerne non

seulement le versement des prestations, mais aussi le placement des chômeurs. De surcroît, les autorités locales arrivaient à contrôler d'assez près les prestataires de l'assurance-chômage. Par exemple, ces derniers étaient obligés de se présenter tous les 15 jours au bureau local. Ceux qui ne s'exécutaient pas perdaient le droit aux prestations.

Figure 2 :

Écart entre les niveaux de chômage selon le RPAC et selon l'EPA

Finlande, de 1961 à 1995 (moyennes annuelles)



En raison des facteurs susmentionnés et, parallèlement, de la croissance économique rapide qui a permis de mettre en place le régime d'assurance-chômage, la plupart de l'offre excédentaire de main-d'oeuvre - autrement dit, les chômeurs - a effectivement été prise en charge par les services publics de placement.

2.2 Rétablissement des statistiques?

Concentrons-nous maintenant sur les huit dernières années, période durant laquelle les données du registre ont produit systématiquement des estimations du chômage un peu plus élevées que les données de l'enquête sur la population active. L'écart entre les deux statistiques a augmenté parce que le contrôle officiel exercé sur les personnes - ici les chômeurs - a diminué et que l'État s'est distancé du rôle de contrôleur et d'administrateur pour adopter celui de fournisseur de services, revirement souvent qualifié de «retrait de l'État» dans le discours politique néolibéral.

Un des indicateurs clés du contrôle qu'exerce l'administrateur sur les prestataires est la fréquence à laquelle ces derniers sont obligés de prendre contact avec l'autorité concernée. Durant les années 70, chaque prestataire de l'assurance-chômage devait se présenter aux autorités compétentes tous les 15 jours. Toutefois, durant les années 80, l'intervalle entre deux enregistrements est passé progressivement de deux semaines à un mois.

À la suite d'une décentralisation, la décision quant à l'intervalle entre deux visites obligatoires des prestataires de l'assurance-chômage au bureau local a été prise non plus par l'administration centrale mais par les bureaux de district des services de placement. À leur tour, ces bureaux de district ont délégué la responsabilité aux bureaux locaux.

À partir de 1988, chaque bureau local des services de placement a eu le droit de décider de la fréquence à laquelle les prestataires devaient se présenter. La plupart des bureaux ont étendu l'intervalle entre les visites d'un mois à deux ou trois mois. Plus tard, les pratiques de déclaration se sont encore relâchées

davantage et, aujourd'hui, la fréquence à laquelle les prestataires doivent se présenter au bureau des services de placement est généralement déterminée individuellement, par l'agent chargé du dossier. L'intervalle entre les déclarations a augmenté ainsi de 6 à 12 mois.

Les contacts peu nombreux des prestataires avec les bureaux du chômage causent des retards d'enregistrement qui sont à l'origine du décalage entre les statistiques établies d'après les données du registre et d'après celles de l'EPA. Ce décalage correspond à la proportion de personnes qui ont réussi à décrocher un emploi mais qui sont toujours officiellement enregistrées à titre de chômeurs. L'écart dépend aussi de la situation générale de l'économie : le décalage est d'autant plus accentué que la conjoncture économique est bonne et que la probabilité de trouver un emploi augmente. En revanche, pendant une récession, le décalage diminue. La figure 3 illustre ces effets : durant la période de croissance économique rapide allant de 1988 à 1990, l'écart entre les mesures a grimpé en flèche, passant de 8 % à 17 %. Quand la récession a débuté en 1991, les deux séries de données se sont de nouveau rapprochées, l'écart étant aujourd'hui de l'ordre de 8 % (voir aussi Koskimäki, 1992).

3. FIABILITÉ D'UN SYSTÈME STATISTIQUE AXÉ SUR UN REGISTRE

Le cadre conceptuel de l'analyse et de la description des statistiques fondées sur les données d'un registre est assez pauvre comparativement aux outils méthodologiques et conceptuels dont on dispose pour évaluer les statistiques fondées sur des données d'enquête. Selon Myrskylä, Tauber et Knott (1995, 321), jusqu'à présent, les études portant sur la qualité des données administratives se sont limitées à l'examen de la cohérence des données chronologiques qui figurent dans la base de données et à des enquêtes par sondage effectuées à titre de contrôle de la qualité.

Dans la description qui suit, je m'appuierai, pour rendre l'analyse plus systématique, sur les concepts courants de fiabilité et de validité. Le terme «fiabilité» désigne tout simplement la capacité qu'a le système statistique de mesurer «ce qu'il est destiné à mesurer» (Brackstone, 1987, 32). Le terme «validité», quant à lui, désigne ici l'utilité du registre en tant qu'indicateur du comportement de la population active.

3.1 Fiabilité technique

Le calcul des statistiques sur le chômage d'après les données du registre des prestataires de l'assurance-chômage est assez simple. Une fois par mois, à la fin de la dernière journée ouvrable, on effectue un dénombrement transversal des «périodes de chômage valides» inscrites dans le registre. Les erreurs éventuelles résultent des pratiques appliquées pour enregistrer les nouvelles périodes de chômage et pour supprimer les «périodes terminées». L'origine des erreurs est facile à dépister aussi : soit l'administrateur enregistre incorrectement les renseignements communiqués par le prestataire, soit celui-ci omet de communiquer des renseignements à l'administrateur.

Dans la suite de l'article, j'appliquerai deux méthodes pour évaluer le rendement du registre. La première se base sur le système d'assurance de la qualité mis en place par l'administrateur du registre. La deuxième se fonde sur des travaux accomplis par le Bureau de la statistique de la Finlande en vue de recalculer le nombre de prestataires de l'assurance-chômage à partir de versions épurées et plus complètes du registre.

L'administrateur du registre effectue des contrôles de qualité chaque année au début de l'année statistique. Ces contrôles consistent à calculer le nombre de prestataires de fin de l'exercice après que le registre ait été mis à jour pendant un mois et de comparer le résultat aux statistiques publiées en se basant sur un registre mis à jour un ou deux jours seulement. Une comparaison est faite aussi entre les données publiées et celles calculées d'après un registre mis à jour pendant 13 mois. Les résultats de ces comparaisons sont présentés à la figure 3.

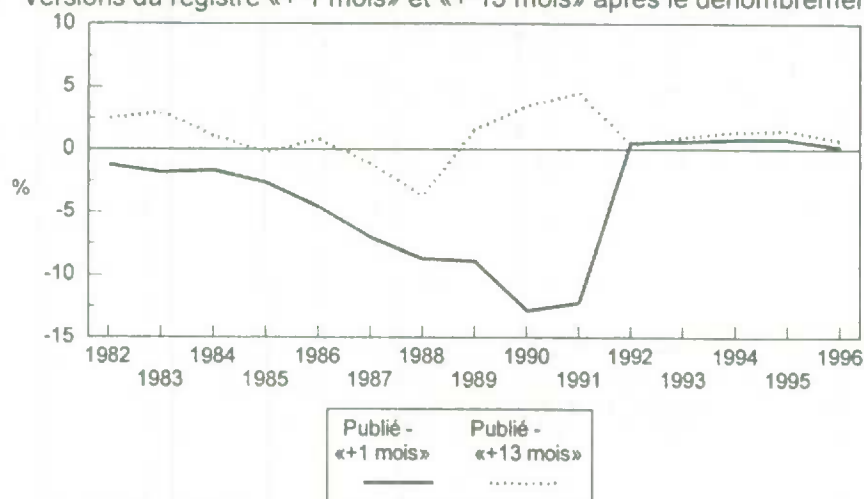
Le comportement du registre change manifestement en 1991. Durant la période antérieure à 1991, les dénombrements obtenus à partir de différentes versions du registre fluctuent : la mise à jour du premier mois correspond en grande partie à des cas supplémentaires de chômage, de sorte que la version «+1 mois» du registre indique jusqu'à 10 % de chômeurs de plus que la version d'après laquelle les données ont été publiées. Durant les 12 mois suivants, les mises à jour portent surtout sur les «fausses périodes valides», de sorte que la version «+13 mois» du registre indique de nouveau un nombre de chômeurs légèrement inférieur au chiffre publié. Après 1991, comme la plupart des mises à jours ont correspondu à la suppression de fausses périodes de chômage, la version «+1 mois» ainsi que la version «+13 mois» du registre indiquent un nombre de chômeurs inférieur au nombre publié.

La deuxième vérification technique consiste à effectuer un dénombrement mensuel d'après des versions du registre mises à jour et contenant aussi des renseignements sur les mises à pied à court terme. Les personnes mises à pied sont comptées comme des chômeurs lors des dénombrements mensuels des prestataires de l'assurance-chômage, mais ne sont pas incluses dans le contrôle de qualité annuel effectué par l'administrateur du registre². Dans le cadre de la présente étude, en plus des contrôles de qualité type, on a vérifié toutes les données mensuelles pour chaque année de référence au lieu de se limiter à la situation de fin d'exercice.

Figure 3 :

Comparaison du dénombrement transversal publié aux chiffres du registre mis à jour :

versions du registre «+ 1 mois» et «+ 13 mois» après le dénombrement



- fin de décembre

La figure 4 illustre les résultats de la deuxième vérification technique. Le dénombrement des chômeurs effectué d'après le registre mis à jour est systématiquement inférieur au dénombrement transversal (sauf pour décembre 1990). Donc, l'effet de la suppression tardive des enregistrements correspondant à des périodes de chômage terminées est plus important que celui de l'enregistrement tardif des nouvelles périodes de chômage. En outre, les résultats varient considérablement d'un mois à l'autre de l'année.

Depuis décembre 1990, l'écart relatif dû à la mise à jour est nettement moindre qu'il ne l'était à la fin des années 80. Avant 1990, le dénombrement transversal des chômeurs est environ 30 % plus élevé que celui effectué d'après le registre mis à jour. Dans une certaine mesure, l'écart dû au retard de mise à jour semble

² Selon le mois et la conjoncture économique, le nombre publié de prestataires de l'assurance-chômage comprend de 5 % à 25 % de personnes mises à pied temporairement.

aussi augmenter durant l'été, puis diminuer en hiver. Après 1990, l'écart n'est plus que de l'ordre de 5 % et on ne décèle aucune variation saisonnière précise.

On peut donner au moins trois explications à la variation du niveau et de l'aspect de la courbe du décalage. Premièrement, la conjoncture économique a changé de façon spectaculaire. Le nombre de chômeurs a quadruplé en deux ans, ce qui a causé un engorgement important dans les bureaux locaux de placement. Il est fort probable que le résultat exceptionnel observé pour décembre 1990, c'est-à-dire dénombrement des chômeurs d'après le registre mis à jour légèrement supérieur au dénombrement transversal, soit dû au fait que les bureaux locaux n'ont pas pu absorber le flux de nouveaux requérants.

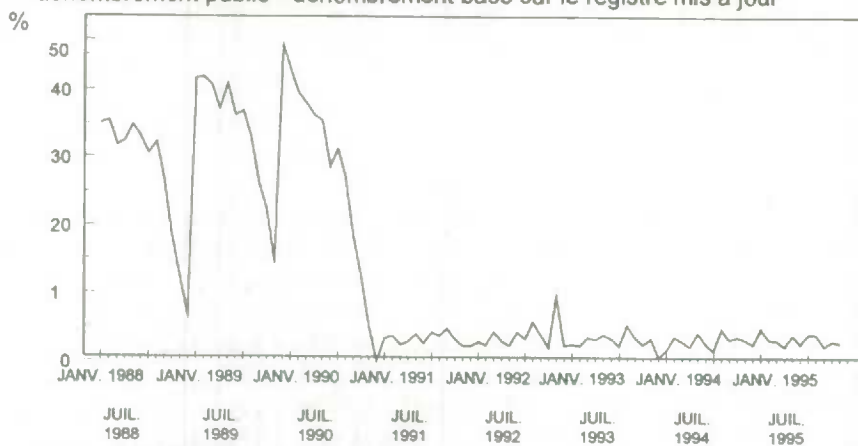
En outre, la possibilité de trouver un emploi étant limitée à cause de la gravité de la récession, le registre a été tenu plus à jour, simplement parce que la proportion de périodes de chômage terminées et d'erreurs d'enregistrement de ces périodes a diminué.

Deuxièmement, le calcul des statistiques est devenu centralisé : avant 1991, le système informatique reposait sur quatorze bases de données administrées par les districts, mais la planification du contenu et de la structure de ces bases de données relevait du bureau central. On calculait les statistiques sur les prestataires de l'assurance-chômage en deux étapes. On effectuait d'abord les dénombrements par district, puis le dénombrement national d'après les chiffres des districts. Depuis 1991, le système informatique comprend une base de données centrale permettant de calculer directement le chiffre national.

L'effet des changements techniques inhérents à ce nouveau système est probablement moins important que celui de l'imposition simultanée aux bureaux locaux et aux bureaux de district de nouvelles normes et règles concernant l'enregistrement des chômeurs. De surcroît, l'établissement d'une base de données centrale a permis aux dirigeants du bureau central de mieux contrôler les activités des bureaux de district. Ces changements semblent avoir eu une incidence cruciale sur la qualité des données du registre. En effet, la qualité administrative de ce dernier, qui était médiocre avant 1991, est devenue tout à fait satisfaisante après la centralisation.

Figure 4 :

Écart d'origine technique entre les statistiques sur les prestataires de l'assurance-chômage
- dénombrement publié - dénombrement basé sur le registre mis à jour



- y compris les mises à pied temporaires

Troisièmement, certaines mesures législatives prises à cause de la récession ont limité la couverture du régime d'assurance-chômage. La décision d'allonger la période d'activité requise pour avoir droit aux prestations d'assurance-chômage explique la modification de la fluctuation saisonnière de l'écart. Durant

les années 80, la période d'emploi obligatoire n'était que de trois mois. En raison de la brièveté de cette période, la plupart des étudiants à la recherche d'un emploi d'été s'inscrivaient à titre de chômeurs.

Nombre de ces étudiants trouvaient un emploi d'été, puis poursuivaient leurs études à l'automne, mais restaient inscrits dans le registre à titre de chômeurs et n'en étaient rayés qu'au moment où ils omettaient de se réinscrire. Durant les années 1990, la période d'activité obligatoire a été portée à six mois et les étudiants à la recherche d'un emploi d'été n'ont plus été acceptés à titre de chômeurs, à moins qu'ils ne déclaraient explicitement ne pas avoir l'intention de continuer leurs études.

3.2 Fiabilité du contenu

Le chapitre précédent décrit les erreurs dues à des lacunes administratives. Un autre ensemble d'erreurs éventuelles est lié au comportement des prestataires lors de leurs visites au bureau des services de placement, ou au fait qu'ils omettent de se présenter. Deux types d'erreurs peuvent se produire ici. L'un survient quand un chômeur ne s'inscrit pas au registre alors qu'il conviendrait qu'il le fasse. Ce type d'erreur est fort semblable aux problèmes de validité examinés au chapitre suivant. L'autre type, plus courant, se produit quand le bureau n'obtient aucun renseignement sur la situation réelle du prestataire. Celui-ci est alors rayé du registre après l'échéance de la période de réenregistrement. Le prestataire pourrait aussi avertir tardivement le bureau du chômage de la modification de sa situation. Selon les pratiques du bureau, la date de fin de période de chômage enregistrée peut correspondre à la date où le renseignement sur la fin de la période est communiqué au bureau ou à la date «réelle» de la fin de la période.

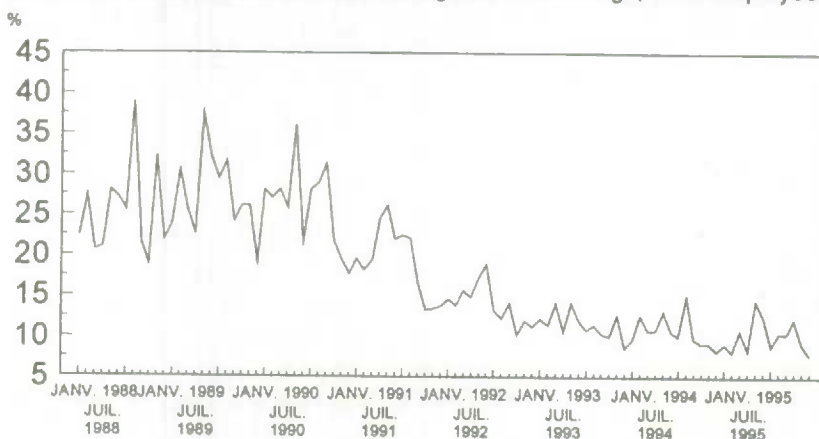
Pour mesurer ce type d'erreur, on a couplé, au niveau individuel, les données du registre qui indiquent si une personne est inscrite ou non à titre de chômeur aux données mensuelles de l'EPA'. Puis, on a effectué un recoupement entre l'état d'activité selon l'EPA et l'état de chômeur à la recherche d'un emploi indiqué par le registre. À titre d'estimation de ce type d'erreur dans le registre, on donne, à la figure 5, la proportion de chômeurs qui, selon les résultats de l'EPA, travaillaient effectivement.

Ici aussi, on note l'effet du cycle économique. Durant les années de croissance, près de 30 % des chômeurs enregistrés travaillaient en réalité, selon l'EPA. Quand la récession a débuté, la proportion de prestataires de l'assurance-chômage ayant effectivement un emploi a baissé pour atteindre environ 10 %.

Pour obtenir une mesure grossière de la fiabilité d'un dénombrement transversal basé sur les données du registre, on peut calculer la somme des erreurs techniques et des erreurs de contenu. On constate ainsi que la fiabilité technique du dénombrement des chômeurs d'après les données du registre, c'est-à-dire le degré d'exactitude de la mesure du chômage selon les définitions administratives, varie considérablement au fil du temps : de l'ordre de 0,4 seulement durant la croissance économique des années 80, elle s'est améliorée durant les années de récession pour s'établir à plus de 0,8.

Figure 5:

Écart entre les statistiques sur le chômage lié au contenu des données
- proportion de personnes inscrites au registre du chômage, mais employées selon l'EPA



4. VALIDITÉ DU REGISTRE EN TANT QU'INDICATEUR DU MARCHÉ DU TRAVAIL

Selon les recommandations de l'OIT sur la statistique du chômage, pour être considérée au chômage, une personne doit effectuer réellement des démarches pour obtenir un emploi, autrement dit rechercher activement du travail. Les deux autres critères à remplir consistent à n'avoir aucun emploi et à déclarer être capable de commencer à travailler dans un délai très court. Les démarches particulières en vue d'obtenir un emploi - prendre contact avec un employeur, prendre contact avec le bureau des services de placement, etc. - devraient avoir eu lieu durant la période de quatre semaines précédant la période de référence (enquêtes de... 1990).

4.1 Démarche en vue d'obtenir un emploi

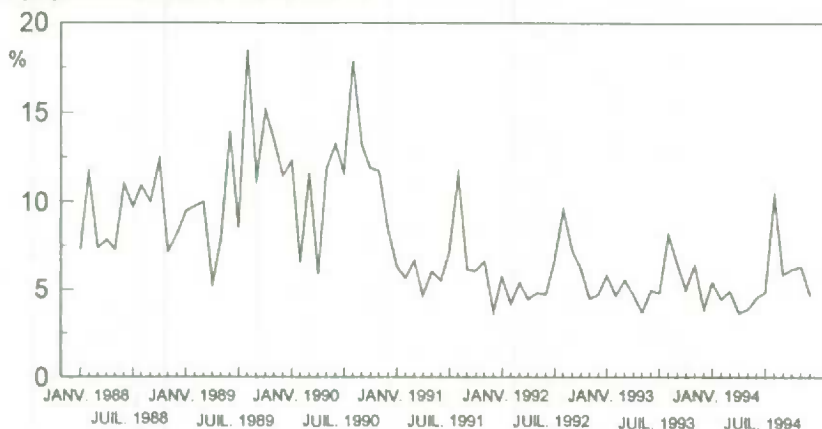
Il est manifeste que les statistiques du ministère du travail incluent des personnes qui ne répondent pas au critère selon lequel des démarches en vue d'obtenir du travail ont été effectuées pendant les quatre semaines précédant la période de référence en tant que tel, car l'intervalle entre les inscriptions obligatoires est beaucoup plus long que les quatre semaines de prospection requises. Toutefois, du point de vue de l'EPA, cela ne signifie pas que tous les prestataires de l'assurance-chômage qui n'ont pas pris contact avec le bureau des services de placement durant la période de quatre semaines requise doivent être exclus de la population de chômeurs. Typiquement, les chômeurs appliquent plusieurs méthodes de recherche d'emploi, de sorte qu'un prestataire de l'assurance-chômage pourrait, dans le cas de l'EPA, être considéré comme activement à la recherche de travail parce qu'il recourt à d'autres méthodes que la consultation des services de placement. La figure 6 montre la proportion de prestataires passifs, c'est-à-dire les personnes inscrites dans le registre comme étant à la recherche d'un emploi, mais qui, selon l'EPA, ne font pas partie de la population active.

La proportion de prestataires de l'assurance-chômage passifs est restée relativement stable, légèrement au-dessous de 10 %, durant la période de référence. La légère diminution observée après 1991 est probablement due à la modification des règlements en ce qui concerne le droit des étudiants à s'inscrire au chômage. La période officielle de congé explique probablement le maximum saisonnier observé en été : durant cette période, il est difficile de chercher du travail et les répondants de l'EPA déclarent

effectivement qu'ils n'ont pas cherché un emploi ou qu'ils ne commenceront à chercher que deux semaines plus tard. Naturellement, ils maintiennent leur inscription au chômage durant cette période.

Figure 6:

Possibilité de rechercher un emploi et activité de recherche d'emploi - proportion de prestataires de l'assurance-chômage ne faisant pas partie de la population active selon l'EPA



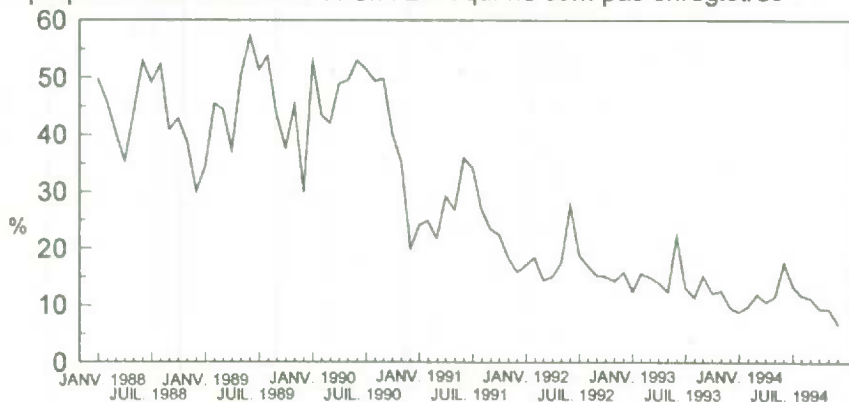
4.2 Couverture du registre

La mesure de la couverture du registre des prestataires de l'assurance-chômage - la proportion de chômeurs selon l'EPA qui ne sont pas inscrits au chômage - est illustrée à la figure 7.

De nouveau, on observe une variation du niveau de couverture qui coïncide avec le cycle économique. À la fin des années 80, environ la moitié seulement des chômeurs selon l'EPA étaient inscrits au registre du chômage; durant les années 90, la proportion de personnes non enregistrées a baissé pour s'établir à un peu plus de 15 %.

Figure 7:

Couverture du RPAC - proportion de chômeurs selon l'EPA qui ne sont pas enregistrés



La plus forte proportion de chômeurs non enregistrés durant la période de croissance économique tient vraisemblablement au fait qu'à cette époque les périodes de chômage avaient tendance à être assez brèves -

il était nettement plus facile d'obtenir un emploi que d'attendre de recevoir des prestations d'assurance-chômage. Par contre, durant les années 90, en raison du renversement de la situation économique, le taux d'inscription au chômage a été très élevé.

Il existe un parallèle entre la tendance saisonnière qui se dégage ici, c'est-à-dire un maximum durant l'été, et l'écart saisonnier dû à la qualité des données du registre observé durant les années 80 (voir la figure 4). La modification des règlements concernant le droit des étudiants à s'inscrire au chômage a fait disparaître l'écart, mais a créé un groupe d'étudiants incapables de trouver un emploi d'été. Comme ces étudiants n'ont plus droit aux prestations d'assurance-chômage, ils ne sont reconnus comme chômeurs que dans le cas de l'EPA.

5. CONCLUSIONS

5.1 Stratégies d'évaluation de la qualité des statistiques administratives

Comme je l'ai mentionné au début, l'évaluation empirique de la qualité des statistiques administratives est souvent difficile. Dans le cadre de la présente étude, on a pu surmonter les problèmes grâce à la possibilité exceptionnelle de coupler le contenu du registre des prestataires de l'assurance-chômage aux données d'une enquête à grande échelle pour une longue période de référence. Ce type d'analyse a également permis de dégager la source la plus importante d'erreur, à savoir la proportion de prestataires de l'assurance-chômage qui, en fait, ont déjà trouvé un emploi.

De surcroît, l'étude montre que la vérification de la cohérence des données au cours du temps est un outil utile d'évaluation de la qualité de ce type de registre.

Le tableau obtenu quand on compare les macrodonnées est assez différent : à titre de macro-indicateurs, les deux sources donnent des résultats à peu près semblables. Cette situation tient, dans une certaine mesure, au fait que les erreurs d'ordre administratif qui entachent le registre se compensent. À l'époque où la couverture du registre était médiocre, les retards de mise à jour étaient exceptionnellement importants. Par conséquent, le taux de chômage calculé d'après les données du registre était le même que celui calculé d'après les données de l'EPA. La présente étude indique notamment que les erreurs qui entachent les statistiques administratives dépendent de la phase du cycle économique. Cette constatation a été faite pour l'EPA dans d'autres cas que celui décrit ici (Eurostat, 1996) et pour d'autres statistiques que celles du chômage, par exemple les statistiques sur la balance des paiements.

Une méthode courante d'évaluation de la qualité des statistiques fondées sur les données d'un registre consiste à insérer une question sur l'inscription au registre/la participation au programme/l'obtention de prestation dans le questionnaire d'une enquête. Cette stratégie est également utilisée pour obtenir le compte national de prestataires quand il n'existe aucun système centralisé permettant d'effectuer le dénombrement total d'après les registres. En théorie, la méthode donne de meilleurs résultats à titre de vérification de la validité qu'en tant que vérification de la fiabilité technique du registre. En effet, l'étude indique que les erreurs significatives sont, pour la plupart, dues à des lacunes administratives au sujet desquelles les répondants d'une enquête ne peuvent fournir aucun renseignement. Les résultats pourraient aussi être entachés d'erreurs propres aux enquêtes¹¹, comme la sous-déclaration dans le cas de questions délicates, dont celles visant le montant des prestations (voir Steel, 1996).

5.2 Utilisation des statistiques administratives

Si on recherche un macro-indicateur général du niveau de chômage, le choix de la série de données importe peu. En revanche, si on veut analyser la série plus en détail, l'instabilité du nombre de prestataires de l'assurance-chômage pourrait poser des problèmes. Le taux de chômage fluctue manifestement (en Finlande) selon la saison et il est nécessaire de désaisonnaliser les données pour effectuer des comparaisons d'un mois à l'autre. L'application de la méthode X11-ARIMA donne de moins bons résultats pour la série de données du registre des prestataires de l'assurance-chômage que pour la série de données de l'EPA, l'erreur qui entache les prévisions faites au moyen du modèle ARIMA étant généralement plus grande et la détection des points de renversement de la tendance, plus difficile.

Le registre des prestataires de l'assurance-chômage a également été utilisé comme base de sondage lors d'études du chômage au niveau micro. Dans ce cas, le risque d'obtenir des résultats erronés et trompeurs à cause des lacunes du registre est plus grand. À la fin des années 80, un échantillon tiré du registre aurait couvert, en prenant comme référence la population de chômeurs selon l'EPA, environ 25 % de la population totale des chômeurs. On court aussi des risques du même genre quand on se sert du contenu du registre - par exemple, les renseignements sur la durée du chômage - pour évaluer la politique de l'emploi ou les travaux sur la dynamique du marché de l'emploi.

L'utilisation des données administratives comme données auxiliaires lors du calcul d'estimations d'après des données d'enquête est considérée par certains comme un effort prometteur en vue d'éviter les erreurs dues et non dues à l'échantillonnage qui entachent les données d'enquête. Bien que cela soit vrai dans une certaine mesure, l'utilisation des données du registre des prestataires de l'assurance-chômage comme données auxiliaires pour le calcul des estimations basées sur l'EPA montre que la variation de la qualité des données du registre se répercute effectivement quand on applique des méthodes telles que la pondération. L'utilisation des données administratives comme données auxiliaires des données d'enquête devrait aussi être examinée soigneusement d'un point de vue qualitatif.

6. BIBLIOGRAPHIE

Alonso, William et Starr, Paul (editors): *The Politics of Numbers. A Census Monographs Series*, Russell Sage Foundation, New York 1987.

Brackstone, G.J. Utilisation des dossiers administratifs à des fins Statistiques, juin 1987, Vol. 13, No. 1, p. 35-51, Statistique Canada.

Davidson, Roger. Official Labour Statistics: a Historical Perspective. *J.R. Statist. Soc. A* (1995) 158, part 1, p. 165-173.

Innes, Judith Eleanor: *Knowledge and Public Policy - The Search for Meaningful Indicators. Second Expanded Edition*. Transaction Publishers, New Brunswick, New Jersey 1990.

Koskimäki, Timo: Rolling Back the Statistics? - The Unemployed in the Employment Service Register and in the Labour Force Survey. *The Yearbook of the Finnish Statistical Association* 1991, p. 9 - 27. Helsinki, 1992 (en finnois avec résumé en anglais).

Myrskylä, Pekka; Taeuber, Cynthia et Knott, Joseph: *Uses of Administrative Records For Statistical Purposes: Finland and the United States. Actes du 1995 Annual Research Conference, 19 au 21 mars, Arlington, Virginia*.

Steel, David. *Options for Producing Monthly Estimates of Unemployment According to the ILO Definition - Report of the work undertaken by the Task Force*. Central Statistical Office, United Kingdom, 1996.

Surveys of Economically Active Population, Employment, Unemployment and Underemployment - An ILO Manual on Concepts and Methods. ILO, Geneva, 1990.

Surveys of Economically Active Population, Employment, Unemployment and underemployment. An ILO manual on Concepts and Methods. International labour Office, Geneva, 1990.

Annexe I :

Chômeurs : dénombrement d'après le RPAC et d'après l'EPA
Finlande, moyennes annuelles, de 1959 à 1995



ⁱ Le couplage a été effectué en se servant du code de sécurité sociale qui est un identificateur unique général utilisé par la plupart des systèmes administratifs de la Finlande. Ce genre de couplage de données n'est autorisé qu'à des fins statistiques.

ⁱⁱ Certaines données de base sur l'EPA mensuelle de la Finlande utilisées ici comme référence : effectif de l'échantillon : environ 12 000 personnes/mois; taux de non-réponse : environ 7 %. Le taux de non-réponse a tendance à être un peu plus élevé pour les prestataires de l'assurance-chômage, particulièrement ceux pour lesquels la période de chômage est longue. Étant donné l'effectif de l'échantillon et le très faible taux de chômage à la fin des années 80, le brouillard qui entache les données chronologiques correspondant à cette période pourrait être dû, en partie, à l'erreur d'échantillonnage.

DISCOURS PRINCIPAL DE CLÔTURE

ERREURS NON DUES À L'ÉCHANTILLONNAGE ET ESTIMATION PAR SONDAGE

Wayne A. Fuller¹

RÉSUMÉ

Les erreurs non dues à l'échantillonnage se retrouvant dans toutes les enquêtes par sondage, le statisticien d'enquête doit adopter une stratégie à plusieurs volets. Durant la conception de l'enquête, le praticien s'efforce d'identifier les sources d'erreurs non dues à l'échantillonnage et de réduire leurs effets au minimum compte tenu des contraintes budgétaires. Les communications du présent symposium illustrent la nature de ces efforts pour diverses catégories d'activités d'enquête. Toutes les données d'enquête étant entachées d'erreurs non dues à l'échantillonnage, il est important de consacrer des ressources à la détermination des propriétés des erreurs de mesure ou (et) d'inclure dans les données collectées des éléments destinés à réduire l'effet des erreurs non dues à l'échantillonnage au moment de l'analyse. On examine ici certaines méthodes d'estimation applicables à des données observées entachées d'erreurs de mesure.

MOTS CLÉS : Erreurs de mesure; quantiles; calage par régression.

Le présent symposium peut être considéré comme une exposition des meilleures pratiques courantes dans le domaine de la méthodologie d'enquête. Claes Anderson a utilisé l'expression *meilleures méthodes courantes*. Selon nous, l'expression signifie non seulement l'utilisation des meilleures procédures connues, mais aussi la conception délibérée d'expériences et d'activités de collecte de données visant à améliorer les procédures actuelles. Groves a donné le ton durant son discours d'ouverture en insistant sur le fait que l'on doit quantifier les effets de l'erreur de mesure en vue d'améliorer les procédures d'enquête. Les exposés de la première séance nous ont rappelé qu'il est important de communiquer les renseignements sur les erreurs de mesure et sur les erreurs d'échantillonnage. Ceux des séances trois, quatre et cinq ont mis l'accent sur les méthodes destinées à réduire les erreurs non dues à l'échantillonnage. Enfin, ceux des séances deux, six, sept et huit ont traité des méthodes d'estimation des propriétés de l'erreur de mesure.

Si plusieurs communications contenaient des exemples de situations où l'observation des propriétés des erreurs de mesure et de non-réponse a mené à la révision des méthodes de collecte ou de traitement des données, rares sont celles qui ont illustré l'application des propriétés des erreurs de mesure au calcul des estimations. Les sujets abordés durant le présent symposium témoignent de ce que l'utilisation de méthodes d'estimation tenant compte de l'erreur de mesure n'est pas courante dans le monde de la statistique d'enquête. Pourtant, le thème même du symposium et ceux d'autres conférences semblables montrent manifestement que les statisticiens d'enquête sont conscients du fait que les données peuvent être entachées d'erreurs. De surcroît, les articles consacrés aux méthodes d'estimation applicables aux données affectées d'une erreur de mesure sont maintenant assez nombreux. À cet égard, consulter Fuller (1987), Carroll, Ruppert et Stefanski (1995) ainsi que les références citées par ces auteurs. Certaines méthodes décrites dans ces articles visent à résoudre des problèmes compliqués et sont donc relativement complexes. Par contre, d'autres sont assez faciles à appliquer compte tenu de la technologie informatique contemporaine.

¹ Wayne A. Fuller, professeur, Department of Statistics, Iowa State University, Ames, Iowa, USA, 50011.

Pour quelles raisons l'application de méthodes de correction de l'erreur de mesure n'est-elle pas courante dans le domaine de la statistique d'enquête? L'une tient au fait que les statisticiens d'enquête procèdent en général selon l'usage établi, autrement dit, s'attribuent des responsabilités précises. Nous sommes nombreux à considérer que notre tâche consiste à concevoir une méthode d'enquête, y compris la définition de la base de sondage, l'établissement du plan d'échantillonnage, la construction d'instruments, la collecte de données et le traitement de ces dernières en vue de produire un ensemble final de données pondérées. Nous décrivons la nature de ces activités dans le rapport que nous rédigeons pour le commanditaire. À l'exception de l'établissement du plan d'échantillonnage, qui est exclu explicitement, ces activités sont celles qui font l'objet du présent symposium. En outre, nous calculons généralement des estimations de la variance d'échantillonnage, ou bien, nous fournissons une méthode pour le faire. Si nous accomplissons bien toutes ces activités, nous sommes tenus en haute estime par nos confrères et consocieurs. Il incombe donc à ceux et à celles qui analysent les données de s'inquiéter de l'effet de l'erreur de mesure sur les résultats.

Il existe cependant d'autres raisons pour lesquelles les statisticiens d'enquête se sont retenus d'appliquer des méthodes d'analyse incluant la correction de l'erreur de mesure. Ces méthodes sont assez complexes et, fait peut-être encore plus important, pratiquement toutes les méthodes d'estimation de l'erreur de mesure nécessitent la définition d'un modèle. Le tirage d'un échantillon aléatoire étant la pierre angulaire de notre discipline, nous hésitons donc à adopter des méthodes dont les résultats se fondent sur les hypothèses d'un modèle. Pourtant, nous ne refusons pas d'appliquer de telles méthodes, comme en témoignent la correction pour la non-réponse, l'imputation pour les données manquantes et l'estimation des petites régions. Les corrections à apporter pour les deux types de données manquantes font maintenant partie des responsabilités acceptées par la plupart des statisticiens d'enquête. Toutefois, la revue d'études anciennes indiquerait sans doute que les méthodes de correction appliquées à la non-réponse n'ont pas toujours été aussi courantes qu'elles ne le sont aujourd'hui. Il fut un temps où, pour de nombreuses enquêtes, on se contentait de mentionner le nombre de non-répondants. L'importance de la non-réponse, et le fait que les statisticiens d'enquête ait déterminé la nature de cette dernière, ont mené à la pratique actuelle consistant à effectuer la correction pour la non-réponse.

Selon moi, les corrections visant à tenir compte de la variance de réponse et du biais dans les réponses finiront par entrer dans la même catégorie que les corrections pour la non-réponse. En fait, les méthodes incluant la correction de l'erreur de mesure sont surtout appliquées quand les mêmes personnes s'occupent de près des opérations d'enquête ainsi que de l'analyse. À cet égard, consultez le numéro spécial de *Statistics in Medicine* (1989), Nusser et coll. (1996) ainsi que les auteurs cités par Carroll, Ruppert et Stefanski (1995).

Dans la suite du présent exposé, je recommanderai certaines méthodes, n'exigeant pas de connaissances très poussées, susceptibles d'être adoptées par les statisticiens d'enquête. Ces méthodes ne sont, certes, pas toujours pleinement efficaces, mais elles permettent d'améliorer les analyses simples, comme les tableaux de fréquence, les régressions et l'estimation de quantile.

Considérons la situation où l'on peut isoler une variable ou un vecteur de variables qui présente un intérêt considérable et dont la mesure est entachée d'erreur. Représentons cette variable ou ce vecteur par la notation X_i ou X_j , respectivement. D'autres variables présentent aussi un intérêt. Les variables de type Y pourraient être des variables dépendantes dans le cas d'une analyse où X est la seule variable explicative. Les variables de type Z sont d'autres variables explicatives susceptibles d'être incluses dans l'analyse. L'âge, la région du pays ou l'urbanisation d'une zone résidentielle sont des exemples de variable Z . Une variable telle que le niveau de scolarité, quant à elle, se situe à la limite entre X et Y . Dans nombre d'analyses, cette variable serait explicative, mais dans certaines, elle pourrait être la variable dépendante.

Supposons qu'on sélectionne un sous-échantillon de l'échantillon original pour étudier l'erreur de mesure. Dans un cas, nous arrivons à déterminer la valeur réelle de x_i dans le sous-échantillon. (Par souci de simplicité, nous considérons le cas d'un scalaire). Supposons de plus, qu'on observe toutes les variables

(x, Y, Z) dans le sous-échantillon. Alors, sous l'hypothèse que le sous-échantillon est un échantillon probabiliste, la configuration est celle d'un échantillonnage à deux degrés classique et la plupart des spécialistes de la statistique d'enquête appliqueraient une des méthodes à deux degrés normalisée. Si le sous-échantillon représente une petite fraction de l'échantillon total et que ce dernier contient beaucoup de variables, on peut toutefois choisir une autre méthode d'estimation fondée sur les techniques décrites plus bas. Grâce à ces dernières, l'analyste peut examiner l'échantillon complet.

Considérons maintenant la situation où le deuxième échantillon n'est pas un sous-échantillon du grand échantillon, mais un échantillon distinct, que nous appellerons «échantillon de calage». En outre, supposons qu'on observe uniquement le vecteur (x, X, Z) dans le sous-échantillon. Dans ce cas, il est possible d'estimer la répartition conjointe de (x, X, Z) en appliquant une méthode à deux degrés, mais il est impossible d'estimer directement la répartition conjointe de (Y, x, X, Z) . Supposons d'abord que x a une densité. Au moyen du sous-échantillon, nous estimons la fonction de répartition cumulative (FRC) de x , représentée par $\hat{F}_x\{\cdot\}$, et la fonction de répartition cumulative de X , représentée par $\hat{F}_X\{\cdot\}$. Supposons aussi que nous calculons la régression de x en fonction de (X, Z) . Autrement dit, nous estimons l'espérance mathématique conditionnelle de x étant donné (X, Z) . Représentons l'espérance mathématique conditionnelle estimée par la relation

$$\hat{x} = g(X, Z, \hat{\theta}),$$

où $\hat{\theta}$ est un vecteur de paramètres estimés. Dans certains cas, g sera un modèle de régression linéaire, mais il pourrait s'agir aussi d'une fonction non linéaire. Afin d'estimer l'espérance mathématique conditionnelle, nous devons pouvoir traiter l'échantillon de calage comme un échantillon probabiliste de la population étudiée.

En se servant des résultats de l'échantillon de calage, on crée un ensemble de données d'analyse tirées du grand échantillon. Le vecteur de variables de l'ensemble de données d'analyse est $(\tilde{x}, \hat{x}, X, Y, Z)$, où

$$\tilde{x}_t = \hat{F}_x^{-1}\{\hat{F}_X(X_t)\} = T(X_t) \quad (1)$$

et $\hat{x}_t = g(X_t, Z_t, \hat{\theta})$. Il existe d'autres moyens de définir $\hat{F}_x^{-1}\{\cdot\}$. Une méthode consiste à définir $\hat{F}_x\{\cdot\}$ comme une fonction continue en joignant les points médians des données de la fonction en escalier empirique FRC par des segments de droite. Plusieurs estimateurs peuvent être construits à partir des données de l'ensemble $(\tilde{x}, \hat{x}, X, Y, Z)$ par des méthodes simples. Il est évident qu'on peut se servir de \tilde{x}_t pour construire des statistiques univariées. Il est manifeste également que les erreurs-types calculées, par exemple, pour les quantiles comme si \tilde{x}_t était un échantillon d'observations, seront trop faibles. Les estimateurs sont des estimateurs à deux degrés et on peut calculer les erreurs-types correctes par des méthodes de répliques ou par d'autres méthodes. À cet égard, consulter Rao et Sitter (1995), Särndal et Swensson (1987), Kott (1990), Breidt et Fuller (1993) et Fuller (1996).

On peut se servir de \hat{x}_t pour construire des estimateurs de la régression de Y en fonction de x et Z . Il n'est pas nécessaire que la régression soit linéaire pour qu'on puisse appliquer les méthodes ordinaires à (Y, \hat{x}, Z) afin d'estimer la régression de Y en fonction de (x, Z) . La méthode exige qu'on émette l'hypothèse selon laquelle l'erreur de mesure est indépendante de l'erreur affectant le modèle. La méthode est applicable, par exemple, au modèle linéaire

$$\begin{aligned} Y_t &= \beta_0 + x_t \beta_1 + e_t \\ X_t &= x_t + u_t, \end{aligned} \quad (2)$$

où la distribution de (e_t, u_t, x_t) a pour moyenne $(0, 0, \mu_x)$ et pour la diagonale de la matrice des covariances $\text{diag}(\sigma_{ee}, \sigma_{uu}, \sigma_{xx})$.

Dans ce cas, la régression de Y_t en fonction de $\hat{x}_t = E\{x_t | X_t\}$ produira un estimateur cohérent de β_1 . Carroll, Ruppert et Stefanski (1995) appellent *calage par régression* la méthode consistant à remplacer x par \hat{x} . Ces auteurs décrivent les trois étapes de la méthode, à savoir :

1. estimer la régression de x en fonction de (X, Z) pour obtenir \hat{x} ,
2. remplacer x par \hat{x} dans l'analyse,
3. corriger les erreurs-types dans l'application de la procédure d'analyse.

La troisième étape indique clairement que les erreurs-types déterminées par l'analyse simple consistant à utiliser \hat{x}_t pour x_t sont incorrectes. Le biais des erreurs-types calculées selon l'analyse standard dépend de la taille de l'échantillon de calage. Si ce dernier représente une petite fraction de l'échantillon de base, l'erreur-type réelle pourrait être considérablement plus importante que celle calculée au moyen des programmes standards à partir de l'échantillon de base. Bien qu'on ne puisse recommander l'utilisation de l'estimateur simple incorrect de la variance, une telle méthode produit habituellement des estimateurs moins biaisés de l'erreur quadratique moyenne que la méthode ordinaire des moindres carrés. Des estimateurs convenables de la variance sont fournis par Carroll, Ruppert et Stefanski (1995) et par Fuller (1987). On peut également recourir aux méthodes des répliques.

Exemple. Considérons une variable X dont le ratio de fiabilité est $\kappa_{xx} = 0,85$ et le modèle de régression

$$Y_t = \beta_0 + \beta_1 x_t + e_t, \quad (3)$$

$$X_t = x_t + u_t,$$

où $(e_t, u_t, x_t) \sim \text{NI}(0, \text{diag}[\sigma_{ee}, \sigma_{uu}, \sigma_{xx}])$, $(\sigma_{ee}, \sigma_{xx}) = (2,00, 0,85)$ et $\beta_1 = 1,00$. Supposons que l'on tire un échantillon aléatoire simple de la population. Alors,

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} \sim \left(\begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} 2,85 & 0,85 \\ 0,85 & 1,00 \end{pmatrix} \right)$$

et $R_{XY}^2 = 0,2535$. Pour des échantillons de taille 1 000, la variance des coefficients calculée par la méthode ordinaire des moindres carrés, $\hat{\gamma}_{1t}$, est

$$V\{\hat{\gamma}_{1t}\} = 0,00213$$

et $E\{\hat{\gamma}_{1t}\} = 0,85$. Si $\kappa_{xx} = \sigma_{XX}^{-1} \sigma_{xx}$ est connu, nous posons

$$\begin{aligned} \ddot{x}_t &= \bar{X} + (0,85)^{1/2} (X_t - \bar{X}) \\ \hat{x}_t &= \bar{X} + 0,85(X_t - \bar{X}). \end{aligned}$$

Alors, la fonction de répartition de \ddot{x}_t dans l'échantillon est un estimateur de la fonction de répartition de x_t . La valeur de \hat{x}_t est l'espérance conditionnelle de x_t étant donné X_t . Les formes simples de \ddot{x}_t et \hat{x}_t ne sont appropriées que si la distribution de (x_t, u_t) est normale. À noter que la variance de \hat{x}_t est plus petite que celle de la population de valeurs réelles.

Si on estime κ_{xx} , c'est-à-dire la corrélation entre les mesures répétées effectuées sur X , à partir d'un échantillon indépendant de taille 100, la variance de $\hat{\kappa}_{xx}$ est, approximativement,

$$V\{\hat{\kappa}_{xx}\} \doteq n^{-1} (1 - \kappa_{xx}^2)^2 = 0,00077.$$

La régression de Y_t en fonction de $0,85X_t$ a une variance de $(0,85)^{-2}(0,00213)$ et une espérance de 1,00 pour un échantillon de taille 1 000. Si on estime le ratio de fiabilité au moyen d'un échantillon

indépendant de taille 100, la variance approximative de l'estimateur de β_1 obtenue par régression de Y_i en fonction de $\hat{x}_i = \bar{X} + \hat{\kappa}_{xx}(X_i - \bar{X})$ est

$$V\{\tilde{\beta}_1\} = V\{\hat{\kappa}_{xx}^{-1}\hat{\gamma}_{1i}\} = (0,85)^{-2}(0,00213 + 0,00077) = 0,00401.$$

Dans ce cas, la variance calculée avec l'estimation de κ_{xx} dépasse de 36 % celle calculée d'après la valeur connue de κ_{xx} . L'erreur quadratique moyenne de l'estimateur ordinaire des moindres carrés est

$$E\{\hat{\gamma}_{1i} - \beta_1\}^2 = (0,15)^2 + 0,00213 = 0,02463$$

Donc, l'estimateur $\tilde{\beta}_1$ est un estimateur essentiellement sans biais dont l'erreur quadratique moyenne vaut un sixième de celle de l'estimateur ordinaire des moindres carrés.

La même méthode générale est applicable quand des variables binomiales, mesurées comme des variables binomiales, sont incluses dans le vecteur X_i . Il n'est pas nécessaire de créer la variable \hat{x}_i pour les variables binomiales, car la distribution marginale de x_i est la même que celle de X_i . La méthode de prédiction par régression exige que l'erreur dans l'équation et l'erreur de mesure de la variable dépendante soient indépendantes de l'erreur de mesure de la variable explicative. Si x_i and X_i sont des variables aléatoires binomiales, il existe une corrélation négative entre u_i et X_i . Cependant, la régression de Y_i en fonction de $E\{x_i | X_i\}$ produit un estimateur cohérent de β_1 si le terme e_i du modèle (2) est indépendant de X_i .

Si les variables dépendante et indépendantes sont des variables aléatoires binomiales, il est nécessaire d'élargir la méthode. Considérons le modèle

$$\begin{aligned} y_i &= \beta_0 + x_i\beta_1 + q_i, \\ Y_i &= y_i + e_i, \\ X_i &= x_i + u_i, \end{aligned}$$

$y_i \sim B_i(1, p_y)$, $Y_i \sim B_i(1, p_Y)$, $x_i \sim B_i(1, p_x)$, $X_i \sim B_i(1, p_X)$, $E\{e_i\} = 0$ et $P\{X = i | x = j\} = \kappa_{ij}$. Comme les variables (Y, X, x) sont situées dans l'intervalle $(0, 1)$, il existe une covariance négative entre u_i and x_i . Il s'ensuit que, en général, il existe une covariance entre e_i and u_i , même si κ_{ij} ne dépend pas de Y_i . Donc, pour la paire (Y_i, X_i) de variables binomiales observées avec erreur, il paraît nécessaire de construire les valeurs espérées pour toutes les configurations possibles. Représentons par Z le vecteur de rangée multidimensionnel à r dimensions défini par les paires (Y, X) ayant les valeurs $[(1, 1), (1, 0), (0, 1), (0, 0)]$. Autrement dit, $Z_i = (1, 0, 0, 0)$ si $(Y_i, X_i) = (1, 1)$ et $Z_i = (0, 1, 0, 0)$ si $(Y_i, X_i) = (1, 0)$. Représentons par z_i le vecteur correspondant défini par les paires (y_i, x_i) . (La variable Y_i peut aussi être mesurée avec erreur.) Étant donné les probabilités de réponse, on peut définir l'espérance de z_i étant donné Z_i .

Considérons maintenant un exemple particulier de l'estimation de la fonction de répartition d'une variable observée sujette à erreur de mesure. Nusser et coll. (1996) décrivent une application particulière de

l'estimation de la distribution des apports habituels d'un constituant du régime alimentaire. Une version simplifiée du modèle empirique utilisé par Nusser et coll. (1996) est

$$\begin{aligned} Y_{ii} &= g^{-1}(X_{ii}), \\ X_{ii} &= x_t + u_{ii}, \\ \begin{pmatrix} x_t \\ u_{ii} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & 0 \\ 0 & \sigma_{uu} \end{pmatrix}\right), \end{aligned} \quad (4)$$

où Y_{ii} représente l'apport observé le jour i pour la $i^{\text{ème}}$ personne. La variable u_{ii} représente la variabilité journalière des apports plus la variance de réponse. Ici, la variance (journalière) d'échantillonnage est considérée comme étant la plus importante. L'apport habituel est donné par

$$y_t = E\{Y_t | x_t\} = E\{g(x_t + u_{ii}) | x_t\}. \quad (5)$$

Représentons la fonction de l'espérance conditionnelle par $h(\quad)$. Alors,

$$y_t = h(x_t). \quad (6)$$

L'équation (6) définit la distribution de y_t , car $x_t \sim N(0, \sigma_{xx})$. En pratique, une seule transformation des observations, $g(Y_t)$, ne suffit pas toujours à produire les observations X_t que l'on peut décomposer en deux composantes normales. Néanmoins, dans le cas des constituants du régime, la transformation des observations originales en valeurs approximativement normales semble produire des données auxquelles on peut appliquer le modèle (4) ou une légère extension de ce dernier.

Dans les études sur la nutrition, les données principales sont les quantités d'aliments que les répondants déclarent consommer quotidiennement. D'après la composition des aliments, on calcule alors l'apport déclaré d'un constituant particulier, comme les protéines. L'objectif consiste à observer, pour une personne donnée, la distribution de l'apport moyen à long terme, appelé *apport habituel*. Ce genre de données est caractérisé par de fortes variations journalières. Environ les deux tiers de la variation observée des apports quotidiens sont dus à la variation propre de la personne et environ le tiers est dû à la variation d'une personne à l'autre. Si on effectue deux observations par personne, environ 50 % de la variation des moyennes individuelles sont dus à la variabilité propre de la personne. Même dans le cas des études où on effectue quatre observations par personne, la variabilité propre à la personne représente le tiers de la variabilité observée des moyennes individuelles.

Selon la méthode de Nusser et coll. (1996), la construction de la transformation g comprend deux étapes. En premier lieu, on choisit une puissance qui rapproche les observations de la normalité. Puis, on se sert d'une fonction spline («en pistolet») pour approximer la transformation de la fonction de répartition des quantiles observés à celle des quantiles de la distribution normale.

On estime le modèle pour les variables transformées par les méthodes d'analyse de la variance. Donc,

$$\begin{aligned} \hat{\sigma}_{uu} &= [n(r-1)]^{-1} \sum_{t=1}^n \sum_{i=1}^r (X_{ii} - \bar{X}_t)^2, \\ \hat{\sigma}_{xx} &= (n-1)^{-1} \sum_{t=1}^n (\bar{X}_t - \bar{X}_{..})^2 - \hat{\sigma}_{uu}, \end{aligned}$$

où

$$\bar{X}_{t.} = r^{-1} \sum_{i=1}^r X_{ti} \text{ and } \bar{X}_{..} = n^{-1} \sum_{i=1}^n \bar{X}_{t.} .$$

Puis, on estime la transformation h pour définir la distribution estimée des apports habituels. En pratique, on estime la distribution pour les divers sous-groupes. (La variable Z définit les sous-groupes.) L'ensemble de données dont a besoin l'utilisateur qui souhaite étudier ces relations en prenant l'apport habituel comme variable explicative doit inclure l'estimation des valeurs espérées. À titre d'exemple, mentionnons l'étude de la relation éventuelle entre l'incidence du cancer et l'apport de lipides.

La figure 1 donne la représentation graphique de l'espérance mathématique estimée de y_t , étant donné (Y_{t1}, Y_{t2}) en fonction de $0,5(Y_{t1} + Y_{t2}) = \bar{Y}_t$ pour l'énergie, pour un ensemble de données sur l'apport en deux jours par personne collectées auprès de 737 femmes de 25 à 50 ans participant à la *Continuing Survey of Food Intakes by Individuals* de 1985 effectuée par le U.S. Department of Agriculture (1987). Les points ne représentent pas une droite, car l'espérance mathématique n'est pas une fonction simple de \bar{Y}_t . Les valeurs représentées graphiquement sont $h(\hat{x}_t)$, où

$$\hat{x}_t = \bar{X}_{..} + (\hat{\sigma}_{uu} + \hat{\sigma}_{xx})^{-1} \sigma_{xx} (\bar{X}_{t.} - \bar{X}_{..}) .$$

La figure 2 montre la représentation graphique des valeurs représentatives du même ensemble de données, où les valeurs représentatives sont

$$T(Y_{t1}, Y_{t2}) = h(\tilde{x}_t) ,$$

et

$$\tilde{x}_t = \bar{X}_{..} + [(\hat{\sigma}_{uu} + \hat{\sigma}_{xx})^{-1} \sigma_{xx}]^{1/2} (\bar{X}_{t.} - \bar{X}_{..}) .$$

La fonction de distribution cumulative des valeurs représentatives fondée sur l'échantillon est un estimateur de la fonction de distribution cumulative des apports habituels. La pente de la droite déterminée par les points représentant l'espérance mathématique estimée est plus faible que celle déterminée par les données correspondant aux valeurs représentatives. Pour l'axe des x , la variance des valeurs représentatives est $\hat{\sigma}_{xx}$, tandis que la variance de l'estimation des valeurs espérées est $(\hat{\sigma}_{uu} + \hat{\sigma}_{xx})^{-1} \hat{\sigma}_{xx}^2$. La représentation graphique de la relation entre l'espérance mathématique estimée et les observations est une courbe, car la distribution originale est asymétrique. Les fonctions obtenues pour d'autres constituants du régime, comme les vitamines, sont encore plus courbes, car la distribution est plus asymétrique.

Carriquiry, Goyeneche et Fuller (1996) ont étendu l'estimation de la distribution de l'apport habituel au cas de la variable bidimensionnelle. Ils emploient une transformation à plusieurs étapes de $Y_{ti} = (Y_{1ti}, Y_{2ti})$ en $X_{ti} = (X_{1ti}, X_{2ti})$ telle que $X_{ti} = g(Y_{ti})$ suive une loi approximativement normale. La figure 3 montre les tracés à 50 % et à 90 % obtenus pour la répartition conjointe estimée de l'apport total de calories et de l'apport de calories provenant des lipides. Les courbes sont tracées pour les observations originales et pour la distribution de l'apport habituel obtenue pour l'échantillon de 4 734 femmes de 20 à 59 ans de la *Continuing Survey of Food Intakes by Individuals* de 1994. Le tracé à 50 % obtenu pour les données originales diffère peu du tracé à 90 % obtenu pour la distribution de l'apport habituel. Les droites qui figurent dans le graphique représentent les moyennes conditionnelles estimées de calories totales, étant donné les calories fournies par les lipides pour les deux distributions. Les deux droites différencieraient davantage si la structure de corrélation de Σ_{uu} était moins semblable à celle de Σ_{xx} .

REMERCIEMENTS

La présente étude a été financée en partie par des fonds résultant de l'entente de recherche n° 58-3198-9-032 et de l'entente de collaboration n° 58-3198-2-006 entre l'*Agricultural Research Service, U.S. Department of Agriculture* et le *Center for Agricultural and Rural Development, Iowa State University*. Je remercie Kevin Dodd et Juan Jose Goyeneche pour l'analyse des données sur la nutrition.

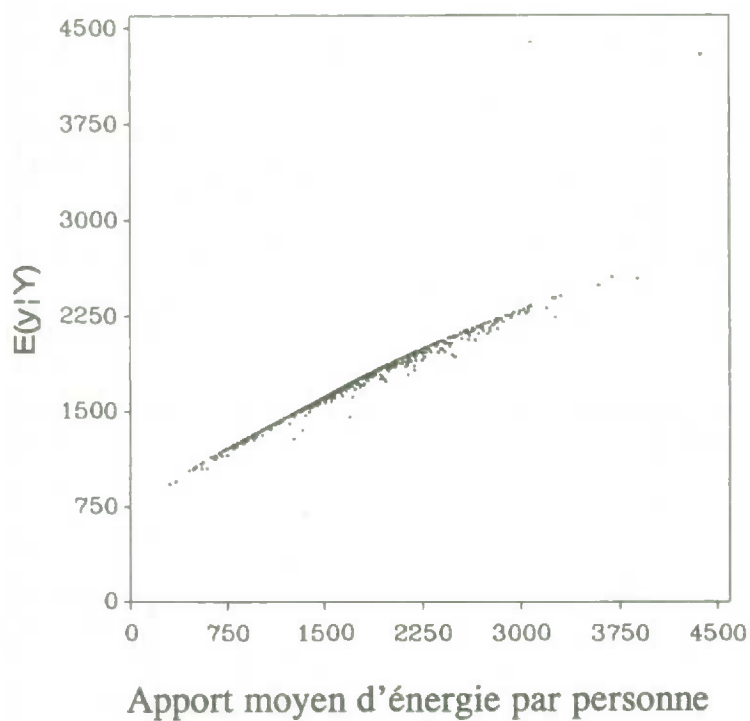


Figure 1. Espérance mathématique estimée de l'apport habituel d'énergie en fonction de l'apport moyen en deux jours.

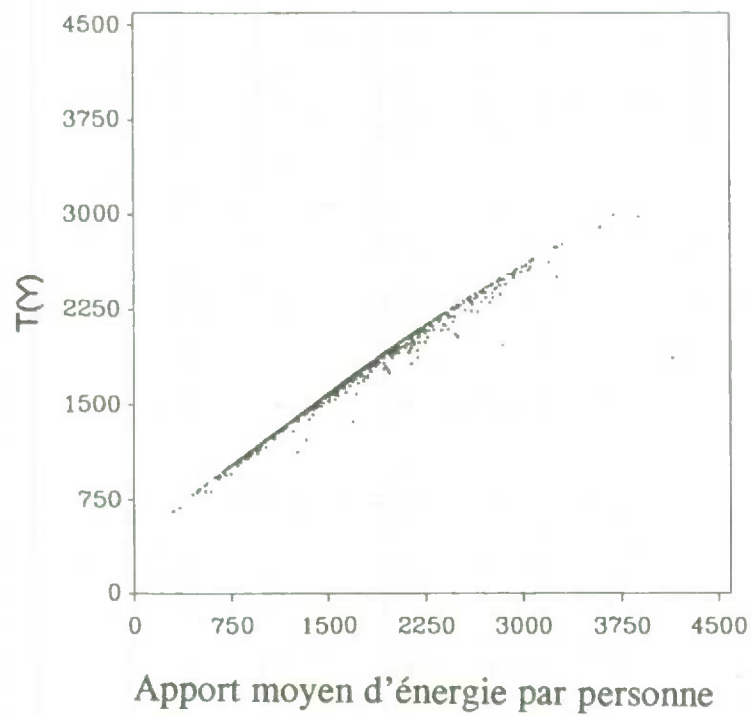


Figure 2. Valeur représentative de l'apport habituel d'énergie en fonction de l'apport moyen en deux jours.

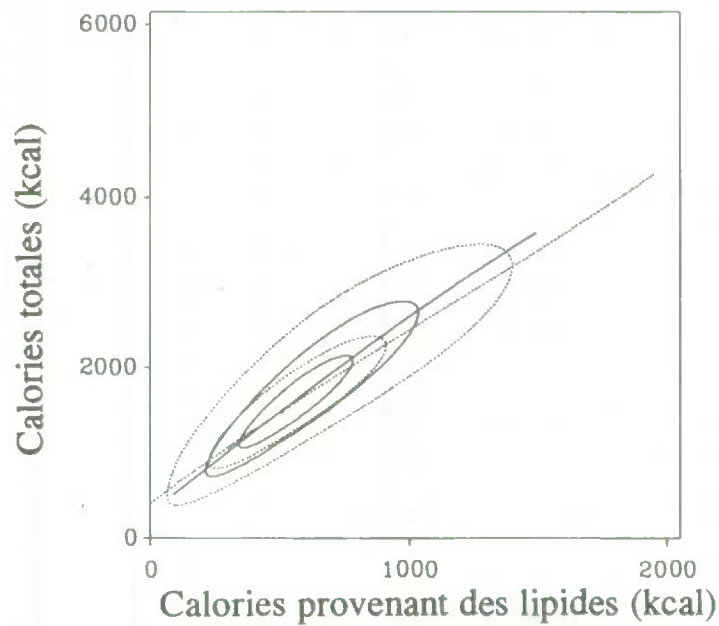


Figure 3. Tracés de la distribution bidimensionnelle de l'apport total de calories et de l'apport de calories provenant des lipides. Les courbes en pointillé représentent les tracés à 50% et à 90% des apports journaliers. Les courbes en trait plein représentent les tracés à 50% et à 90% des apports habituels.

BIBLIOGRAPHIE

- Breidt, F. J. et Fuller, W. A. (1993). Regression weighting for multiphase samples. *Sankhyā Services B* **55**, 297-309.
- Carriquiry, A. L., Goyeneche, J. J. et Fuller, W. A. (1996). Estimation of bivariate usual intake distributions. Rapport non publié. Iowa State University, Ames, Iowa.
- Carrol, R. J., Ruppert, D., et Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, Chapman and Hall, New York.
- Chua, T. C. et Fuller, W. A. (1987). A model for multinomial response error applied to labor flows. *J. Amer. Statist. Assoc.* **82**, 46-51.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd edition. John Wiley, New York.
- Fay, R. E. (1989). Theory and application of replicate weighting for variance calculations. *Proc. ASA Section on Survey Research Methods*, Washington, D.C., 212-217.
- Fuller, W. A. (1996). Replicate variance estimation for two-phase sample. Manuscrit non publié. Iowa State University, Ames, Iowa.
- Kott, P. S. (1990). Estimation de la variance lorsque l'échantillon aréolaire de première phase est restratifié.. *Technique d'enquête* **16**, 107-111.
- Lin, L. I.-K. et Vonesh, E. F. (1989). An empirical nonlinear data-fitting approach for transforming data to normality. *The American Statistician*, **43**, 237-243.
- Mendelsohn, J. et Rice, J. (1982). Deconvolution of Microfluorometric histograms with B -splines. *J. Amer. Statist. Assoc.* **77**, 748-753.
- Nusser, S. M., Carriquiry, A. L., Dodd, K. W., et Fuller, W. A. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*. À paraître.
- Nusser, S. M., Fuller, W. A. et Guenther, P. M. (1996). Estimating usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. In *Survey Measurement and Process Quality*, eds. L. Lyberg, M. Collins, E. DeLeeuw, C. Dippo, W. Schwartz, et D. Trewn. John Wiley, New York.
- Rao, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Off. Statist.* **10**, 153-165.
- Rao, J. N. K. et Shao, J. (1996). On balanced half-sample variance estimation in stratified random sampling. *J. Amer. Statist. Assoc.* À paraître.
- Rao, J. N. K. et Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika* **82**, 453-460.
- Särndal, C. E. et Swensson, B. (1987). A general view of estimation for two-phases or selection with application to two-phase sampling and nonresponse, *Int'l. Statist. Review* **55**, 279-294.

Särndal, C. E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters* **9**, 229-235.

Stefanski, L. A. et Bay, J. M. (1996). Simulation extrapolation deconvolution of finite population cumulative distribution function estimators. *Biometrika*. À paraître.

Stefanski, L. A. et Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169-184.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.

ALLOCUTION DE CLÔTURE

ALLOCUTION DE CLÔTURE

David Binder¹

J'aimerais ajouter quelques observations personnelles au sujet des erreurs non dues à l'échantillonnage. Tout d'abord, qu'est-ce qu'une *erreur non due à l'échantillonnage* par rapport à une *erreur due à l'échantillonnage*. Peut-être que la terminologie laisse à désirer. Après tout, lorsqu'on parle des *erreurs dues à la réponse* par rapport aux *erreurs dues à la non-réponse*, on parle des erreurs qui sont dues aux personnes qui ne répondent pas. Pourquoi donc, dans ce cas, les erreurs non dues à l'échantillonnage ne seraient-elles pas des erreurs dues au fait de ne pas figurer dans l'échantillon? Pour les non-statisticiens, cela pourrait sembler pertinent. Lorsque l'on consulte un ouvrage type sur la théorie de l'échantillonnage, on commence par la définition d'une population comportant un nombre fini d'unités, on choisit un échantillon, puis on discute de l'estimation et d'autres aspects de la stratégie d'échantillonnage. On explique alors que tout écart entre la valeur réelle et nos estimateurs correspond à l'erreur d'échantillonnage. Nous, les statisticiens, pensons que les erreurs non dues à l'échantillonnage sont tout le reste. Toutefois, est-ce que toutes les erreurs, sauf celles dues de façon particulière à l'échantillonnage, sont des erreurs non dues à l'échantillonnage? Par exemple, lorsque vous conduisez et que vous effectuez un mauvais virage à gauche et que vous avez un accident, s'agit-il là d'une erreur non due à l'échantillonnage? Il nous faut donc des définitions plus claires et compréhensibles.

Le domaine des erreurs non dues à l'échantillonnage, même dans un sens strictement statistique, est vaste, et nous avons tendance à le définir comme une série de domaines secondaires. Nous parlons d'erreurs dues à la base sondage, d'erreurs dues à la non-réponse, d'erreurs dues à la réponse, d'erreurs de traitement, etc. Le fait est que la plupart des processus utilisés dans le cadre des enquêtes comportent une intervention humaine et que celle-ci est sujette à l'erreur. Au cours du présent symposium, nous avons discuté principalement des erreurs dues à l'intervention humaine. C'est pourquoi, il semble souhaitable de se tourner vers des experts du comportement humain. Les processus cognitifs de l'être humain peuvent aider à mieux comprendre les causes des erreurs non dues à l'échantillonnage.

En 1978, Statistique Canada publiait un ouvrage intitulé «Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes». On peut se demander dans quelle mesure nous avons progressé au cours des vingt dernières années. Lorsque l'on examine les sujets principaux compris dans la table des matières, on note qu'après l'introduction, l'ouvrage porte sur la couverture, la réponse, la non-réponse, le codage, la saisie des données, la vérification et l'imputation, et d'autres sujets. Il est intéressant de noter que nous connaissions déjà ces problèmes en 1978, et que nous avions déjà une idée de la façon de mesurer et d'évaluer ce genre d'erreurs. Toutefois, lorsque l'on compare le contenu de l'ouvrage et les sujets abordés dans le cadre du présent symposium, il est clair que nous avons fait des progrès substantiels. Mais nous avons encore beaucoup à apprendre.

Il est ressorti très clairement au cours des discussions qu'il existe une interaction très grande entre la gestion des enquêtes, le processus de mesure et d'évaluation des erreurs et la rétroaction nécessaire. Souvent, dans le cadre du présent symposium, on s'est interrogé au sujet des avantages de la mesure et de l'évaluation des erreurs non dues à l'échantillonnage. Toutefois, le problème vient du fait que nous ne nous posons pas beaucoup de questions au sujet des erreurs non dues à l'échantillonnage tant que les choses vont bien.

Mentionnons par exemple l'Enquête sur le tabagisme au Canada qu'a mentionnée Lecily Hunter, dont les données semblent aller à l'encontre de la perception intuitive des experts, ce qui a soulevé des questions au sujet des sources d'erreurs non dues à l'échantillonnage. Nous devons être sensibles à ces besoins, particulièrement en cette période de restrictions budgétaires. Il est trop facile de passer outre la nécessité de mesurer ces erreurs, mais dans le cas

¹David Binder, Directeur, Division des méthodes d'enquêtes-entreprises, Statistique Canada, Parc Tunney, Ottawa, Ontario, Canada K1A 0T6.



où les estimations semblent contestables, nous devons pouvoir répondre aux préoccupations des utilisateurs.

Récemment, on a beaucoup parlé dans la presse des ratés des sondages pré-électorales relatifs à la course à la présidence des États-Unis. En effet, l'écart était beaucoup plus grand entre Clinton et Dole dans les sondages qu'au moment du scrutin, et on s'est posé des questions sur ce qui a pu entraîner cette situation. Non seulement s'est-on interrogé sur la pertinence de l'échantillonnage non probabiliste, mais aussi sur les diverses sources d'erreurs d'échantillonnage. Par ailleurs, dans le cas des sondages qui se sont tenus avant le référendum québécois, en octobre 1995, leurs résultats ont été presque identiques à ceux du scrutin. Et on ne s'est pas posé de question à ce sujet. Il semble qu'on se préoccupe beaucoup des erreurs non dues à l'échantillonnage lorsque ces erreurs sont importantes.

Il convient en outre de souligner la question de la mesure et de l'ajustement pour le sous-dénombrement dans le recensement. Si la couverture de l'enquête semble trop limitée, il est tout à fait justifié de rajuster les nombres, étant donné que les erreurs non dues à l'échantillonnage sont excessives. Par ailleurs, si les erreurs sont minimales, on peut penser que le rajustement en fonction des erreurs non dues à l'échantillonnage occasionne d'autres. Force est de constater que ces questions sont assez complexes et que je n'ai fait ressortir qu'un aspect.

On pourrait citer de nombreux autres exemples. Les gestionnaires d'enquête doivent évaluer avec soin les avantages de la mesure des erreurs non dues à l'échantillonnage par rapport à son coût, en tenant compte des risques d'obtenir des données douteuses.

En terminant, j'aimerais remercier les organisateurs du présent symposium pour la qualité des discussions. Au total, environ 30 personnes ont participé à un aspect ou l'autre de l'organisation du présent symposium. Le principal comité organisateur, dirigé par Éric Rancourt, comprenait Ann Brown, Jane Burgess et Johane Dufour. On a aussi pu compter sur un certain nombre de sous-comités. Parmi les présidents des sous-comités, on comptait Michelle Simard, qui s'est occupée de la logistique, Josée Morel, qui a organisé les discussions en table ronde, Hew Gough, qui s'est occupé des ateliers, et Lynn Savage ainsi que Micheline Sabourin, qui ont assuré la majeure partie du soutien au cours des derniers mois. La correspondance a été principalement prise en charge par Sophie Dionne et Guylaine Dubreuil.

L'an prochain, le symposium s'intitulera «Nouvelles orientations pour les enquêtes et les recensements». Jack Gambino est le président du comité organisateur.

Enfin, j'aimerais encore une fois remercier tous les participants, les présidents de session et toutes les personnes qui ont contribué aux discussions et qui ont fait de ce symposium un succès.