SURVEY METHODOLOGY

TECHNIQUES D'ENQUÊTE

December - 1979 - Décembre

VOLUME 5

NUMBER 2 - NUMÉRO 2

A Journal produced by
Methodology Staff
Statistics Canada

Préparé par les
méthodologistes de
Statistique Canada

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE

December/décembre 1979 Vol. 5 No. 2

A Journal produced by Methodology Staff, Statistics Canada.

Preparé par les méthodologistes de Statistique Canada.

CONTENTS

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE

December/décembre 1979 Vol. 5 No. 2

A Journal produced by Methodology Staff, Statistics Canada.

Preparé par les méthodologistes de Statistique Canada.

Editorial Policy:

The objective of the Survey Methodology Journal is to provide a forum in a Canadian context for publication of articles on the practical applications of the many aspects of survey methodology. The Survey Methodology Journal will publish articles dealing with all phases of methodological development in surveys, such as, design problems in the context of practical constraints, data collection techniques and their effect on survey results, non-sampling errors, sampling systems development and application, statistical analysis, interpretation, evaluation and interrelationships among all of these survey phases. The emphasis will be on the development strategy and evaluation of specific survey methodologies as applied to actual surveys. All papers will be refereed; however, the authors retain full responsibility for the contents of their papers and opinions expressed are not necessarily those of the Editorial Board or the Department. Copies of papers in either Official Language will be made available upon request.

Politique de la rédaction:

La revue Techniques d'enquête veut donner aux personnes qu'intéressent les aspects pratiques de la conduite d'enquêtes, la possibilité de publier sur ce sujet dans un cadre canadien. Les Textes pourront porter sur toutes les phases de l'élaboration de méthodes d'enquête: les problèmes de conception causé par des restrictions pratiques, les techniques de collecte de données et leur incidence sur les résultats, les erreurs d'observation, l'élaboration et l'application de systèmes d'échantillonnage, l'analyse statistique, l'interprétation, l'évaluation et les liens entre les différentes phases d'une enquête. On s'attachera principalement aux techniques d'élaboration et à l'évaluation de certaines méthodologies appliquées aux enquêtes existantes. Tous les articles seront soumis à une critique, mais les auteurs demeurent responsables du contenu de leur texte et les opinions émises dans la revue ne seront pas nécessairement celles du comité de rédaction ni de Statistique Canada. On pourra se procurer sur demande des exemplaires d'un article dans l'une or l'autre langue officielle.

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE

December/décembre 1979 Vol. 5 No. 2

A Journal produced by Methodology Staff, Statistics Canada.

Preparé par les méthodologistes de Statistique Canada.

Submission of Papers:

The Journal will be issued twice a year. Authors are invited to submit their papers, in either of the two Official Languages, to the Editor, Dr. M.P. Singh, Census and Household Survey Methods Division, Statistics Canada, 6th Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6. Two copies of each paper, typed space-and-a-half, are requested. Authors of articles for this journal are free to have their articles published in other statistical journals.

Présentation de documents pour publication:

La revue sera publiée deux fois l'an. Les auteurs désirant faire paraître un article sont invités à en faire parvenir le texte au rédacteur en chef, M. M.P. Singh, Division des méthodes d'enquêtes ménages et du recensement, Statistique Canada, 6e étage, Ediface Jean Talon, Tunney's Pasture, Ottawa, Ontario, K1A 0T6. Prière d'envoyer deux exemplaires, dactylographiés à interligne et demi. Les auteurs des articles publiés dans cette revue sont libre de les faire paraître dans d'autres revues statistiques.

DATA, STATISTICS, INFORMATION - SOME ISSUES OF THE CANADIAN SOCIAL STATISTICS SCENE[1]

Ivan P. Fellegi[2]

This paper looks at the current state of development of social statistics in Canada. Some key concepts related to statistics and social information are defined and discussed. The availability and analysis of administrative data is highlighted, along with the need for social surveys. Suggestions are made about the types of data analysis needed for the development of social decision models to meet policy requirements. Finally, an outline of priorities for future work toward the effective use of social statistics is given.

1. INTRODUCTION

I should start by apologizing for my temerity to address an audience of demographers and sociologists on a topic about which I know so little. Perhaps, I should seek solace in the old Hungarian proverb which, roughly translated, says that "If God gives power to someone, He will be good enough to give some brains to go with it". At any rate, in my previous position at Statistics Canada, I was quite successful at being the best computer expert among survey designers, and vice versa. Perhaps, I can pull off some similar tricks in my new position - starting with this talk.

I have just spent nine months in the United States on a Presidential reorganization project to review the U.S. Federal Statistical System. Particularly, in light of my new responsibility, I could not help observing with some envy, the social statistics data base available there. At least in terms of subject coverage, social statistics in

---

1 A verbatim reproduction of a paper presented at the Learned Societies Conference, Saskatoon, June 2, 1979.

2 Ivan P. Fellegi, Assistant Chief Statistician, Social Statistics Field, Statistics Canada.

the U.S. are certainly much more developed than in Canada. It is tempting to speculate why.

You and I, after perhaps a few minutes of blaming one another, could probably quickly agree on blaming "the powers that be" for allocating insufficient resources to social statistics. Before we reach that inevitable conclusion, however, perhaps we can pause for a few minutes and ask ourselves: why should they invest more resources in social statistics? Suppose they invested considerable funds in social statistics; would they become better informed with respect to issues they can influence through social policies?

2. SOME KEY CONCEPTS

By and large, what public policy discussion needs is insight and information. We tend to offer statistics as a substitute. I think the difference between these notions is of key importance in trying to understand the present state of social statistics in Canada. Let me share with you my attempt to clarify in my own mind the difference between these and a few other related concepts. Those of you having a preference for visual presentations may trace an approximation of these concepts on Chart 1 at the end of the paper.

1. Datum: A datum is a quantity (e.g., the dollar value of sales) or a code (e.g., the numerical code identifying an industry, race, sex, occupation, etc.) which arises out of observation or measurement. A datum must have at least three components: a quantity or code value; a concept[1] which is quantified in the form of a code or quantitative measurement (e.g., the particular way "unemployment" is defined in the Labour Force Survey); and a

---

1 Concepts are abstractions which can be made operational only by selecting some dimension of the "real world" as a measureable proxy to represent that concept.

reference entity, i.e. an entity to which the quantity or code refers. A reference entity may be a person, business, institution, etc. or it may be a group of such entities (e.g., all persons with a permanent residence in Nova Scotia as of January 15, 1978 and whose ages were between 14 and 21 years). A datum for a given reference entity can, of course, include the quantified measure or code of more than one concept - in which case we can talk about multi-dimensional data.

2. Statistic: A statistic is a summarization of data referring to a unique group of persons, businesses, events, phenomena, etc. The expression "unique" means that the members of the group are unambiguously identifiable: producers and users of statistics can apply the same test to a potential member of the group and come to the same conclusion as to whether it is or is not a member of the group. The group definition may take the form of a list of its members, or it may be in terms of their attributes (e.g., all residents of Nova Scotia on a given date). Thus, a person who does not know the identification of all members of the group (in the form of a list) can nevertheless determine, on the basis of its relevant attributes, whether a given entity (person, business, etc.) is or is not a member of the given group.

   A well-defined group can serve as a reference entity and a statistic can therefore be considered as a datum with that group as its reference entity. When we want to distinguish between a statistic and the data which it summarizes, we refer to the component data as micro data.

3. Interpretable message: A datum or statistic which has been communicated by a person or institution is defined as an interpretable message. For example, when Statistics Canada publishes the proportion of unemployed within a given age-sex group in Nova Scotia, this becomes an interpretable message. Before its publication,

the same rate would be a datum, in fact in this case a statistic. To quote a more special example, data we may beam into outer space are interpretable messages: we do not know whether they are received and interpreted by anyone.

4. Information: The term is derived from the verb "inform". To inform is a process. Information, for purposes of this paper, is defined as the process of conveying an interpretable message as a result of which the receiver of the message acquires knowledge, i.e., becomes better informed. Hence, information involves interpretation.

   An interpretable message has the potential of informing, but it needs a receiver. More precisely, an interpretable message becomes information if it is received by an intelligent receiver who interprets it; i.e., does not screen it out but rather stores it in his/her mind for the purpose of some expected future use. An intelligent receiver means a person with the knowledge needed to "decode" a statistic into the three components of data identified earlier and having the ability to relate the decoded statistic to other aspects of his/her knowledge or experience.

   Whether or not an interpretable message, when received, is screened out or stored can be influenced by the sender: through the medium and presentation used, through repetition, etc. More importantly, the sender can induce storage and interpretation of the message by calling the attention of the receiver to the fact that the message has some intrinsic interest or utility for the receiver.

5. Decision model: I am clearly leading up to the point that an interpretable message, duly sent and received, will generally become information (i.e., stored in the mind of the receiver) if the receiver judges the message to be of some interest, relevance or usefulness for him. Leaving aside the question of

interest, the message can be useful to a person if it relates to a phenomenon about which he/she wishes to make a judgement or decision and the result of that decision has some concrete utility to the decision maker.

Decisions typically have as their objectives the modification of the real world in some fashion, or they represent alternative strategies to respond most effectively to different conditions of the real world. In order to do this most effectively, the decision maker should ideally have some reasonably clear objectives and, further, some utility function whose maximization represents a reasonable trade-off between the costs and benefits of alternative decisions. He is interested in those factors which have a relationship with or impact on his utility, depending on decision or judgement to be made. The consideration of these relevant factors takes place within a formal or informal framework, or thought process that will be referred to as a decision model. Before going on, I want to emphasize that the decision model as understood here need not be a formalized mathematical or probabilistic model. It can range all the way from complex econometric or simulation models to an unstructured accumulation of experience. In fact, even where complex and formal models exist, few major decisions are made automatically on the basis of the model's predictions: typically these are tempered by judgements. My definition of a model, for purposes of the present discussion, therefore includes the entire mental process involved in decision making.

Consider three examples. The first example involves a farmer's decision to apply X tons of fertilizer to a corn field so as to maximize net profits. The utility function can be quite well defined as a trade-off of costs and benefits to achieve the single objective. Data on the cost and performance characteristics of fertilizers are directly relevant. The second example

might involve a decision by a person to spend some amount of money either on a vacation or on improving the resale value of his house. The value to the person of the alternative decisions would be very difficult or artificial to incorporate into a formal model. Yet people do take such decisions, so a decision model is involved. Part of this model may be more precisely formulated: for example, the trade-offs involved in the length, location and level of luxury involved in alternative vacation plans. Clearly, data about travel costs are relevant. A third example, involving a more ambitious social decision, may relate to the objective of reducing the inequalities in the income distribution of Canadians. Here the utility function would be quite difficult to define, partly because the same decision makers are likely to be involved with other objectives and hence ideally the combined utility of all these objectives should guide the decisions, partly because even the single desirable outcome is not sufficiently well understood in terms of causalities. What are the factors affecting income inequality - government transfer payments, tax policies, general state of the economy, education, family backgrounds? In the case of factors thought to be relevant, what is their current state?

The real world is infinitely complex - even the limited segment of it which the decision maker wishes to modify or react to. Necessarily, whehter he is aware of it or not, he can only cope with this complexity within the framework of a simplifying model which, over time, he may be able to elaborate further. The model helps in sorting out the relevant factors that should be considered, i.e. those expected to have some predictable relationship with the outcome which he wants to modify. Further, the model would indicate how these relevant factors interrelate with one another and with the outcome, which of the factors can be modified, what is the likely effect on the outcome of modifying one or more of

the factors, what is the current state of the relevant factors. Note that only the last of these questions can possibly be answered by statistical data, and only the past interrelationship between some of the factors can be answered through statistical analysis.

So, facing the real world, the decision maker needs the simplification offered by his formally articulated or informal model. Another complexity where his model helps the decision maker is known as information overload - the reception of a great variety of data, only some of which is related to his decision problem. The model would indicate, in effect, which data he should be interested in and which can he afford to screen out. From the point of view of the sender of an interpretable message, this role of the decision-maker's model is particularly important: if the sender knows that his message is about a factor which is part of the decision-maker's model, he can be reasonably sure that his message will be interpreted. Further, if the sender is able to articulate a relationship between his message and some factor of known interest to the receiver, then in effect, the sender can cause the decision-maker to evaluate and possibly change his model by incorporating the new factor and thereby render the current message relevant for the receiver thus ensuring that it becomes information.

6. Validity: Data collection typically involves compromises between the concept a decision maker might wish to measure (the "ideal concept") and what is possible and practical to measure (the "operationalized concept"). One may have an ideal concept in mind as to how unemployment status should be defined in the context of a decision problem at hand. Different users faced with different decision problems may well lead to different ideal concepts. However, those involved in actually conducting a household survey may decide that

a concept, in order to be measurable with reasonable accuracy, must be related to some concrete activities of individuals which, if they are questioned about them, they are likely to remember. For example, this consideration was a significant reason for the development of the activity-based concept of unemployment used by the Labour Force Survey (LFS) over a long period of time. The respondent not only has to be able to remember his answers, he should also be disposed to respond, willing to accept the burden of response, etc. All of these considerations may lead a survey taker to accept compromises in the concept to be measured. The distance between a given users's "ideal" concept and the operationalized concept actually used measures the validity of the data for the given use. For example, the operationalized concept of unemployment used in the LFS is not ideal for the purpose of monitoring the number of persons suffering economic hardship as a result of unemployment, thus affecting the validity of the LFS for this purpose.

It is critical to understand that, if the resulting data are to have required validity for a decision maker, the underlying concept must be a close enough approximation within his decision model, of an aspect of the real world. Thus "ideal" concepts arise within a decision context. The unemployment concept mentioned above is decisively affected by the decision context in which this concept was defined - monitoring the labour market as opposed to the meaning of social or economic well-being. This has clearly major statistical policy implications: concepts have to be updated either as a result of changes in the real world, or changes in the decision problems addressed. Furthermore, often a single concept related to a particular phenomenon cannot fit exactly the needs of important but different decision problems; in such cases, the job of a statistical agency is either to find the best available compromise, or to collect (as, for example, in the case of unemployment) sufficient detail to permit the construction of estimates for alternative definitions of the concept. Given resource constraints, the latter alternative is only rarely feasible.

7. Relevance: As indicated above, data only have the potential of becoming information. If the utilization of data by a decision maker would reduce the uncertainty associated with his decision, we say that the data are of relevance to him. Clearly, relevance is a property of the data in relation to a class of users or uses, not a property of the data alone. It is a very broad concept. A decision maker with a well articulated decision problem may, for example, need data on the unemployed. In this case, data on the unemployed are quite likely to be of relevance to him, essentially depending upon the distance between the particular concept of unemployed he needs and the one that is available. Thus, in this particular case, relevance becomes synonymous with validity. However, relevance is the broader concept. A decision maker concerned with "general well-being" might consider data on health, income, housing, cultural activities, etc. all relevant, depending on the operationalized concepts used. A statistical agency wishing to render its data as widely relevant as possible must therefore acquire considerable knowledge of the decision issues and models of its users, as well as skills in operationalizing the concepts most useful to decision makers. Such knowledge is acquired - except in the case where data are collected by the end users themselves - through a variety of analytical activities shedding light on the end users' decision problems, at the very least by maintaining close dialogues with a wide cross-section of end users. Once again, the notion of relevance has major policy implications for a statistical agency.

A necessary (although not sufficient) condition for data to be relevant in a decision context is their explanatory power or relatedness with respect to the object of the decision. In the case of micro data, the inclusion of more than one (carefully chosen) variable can often exponentially increase the explanatory power (relevance) of the retrievable statistics. Data on the distribution of the unemployed by age are clearly vastly more relevant

for most purposes than separate data on the age distribution of the population plus the number of unemployed. Thus, the potential relevance of a micro data base is strongly affected not only by the choice of the concepts measured but also by the richness of the data base. Furthermore, given the fact that most models have to use data from several sources, the useability of a given datum in a model strongly depends on the ease with which it can be used jointly with other data. Thus, another prerequisite for increasing the relevance of data emerges: standardization of concepts.

8. Accuracy: The accuracy of data, broadly defined, is the extent to which the actual measurement of the operationalized concept any hypothetical, error-free counterpart are close to one another. It includes the well known components of measurement and, when applicable, sampling errors.

   Furthermore, accuracy is also affected by the extent to which the reference entity, in this case a group, is incorrectly identified, for example, by failing to include in the group, persons who according to the group definition, belong to it. Accuracy which is inadequate for a particular application may render data irrelevant. Put differently, accuracy commensurate with a given substantive objective is one of the many attributes of relevant data.

9. Misleading data: The notions of validity and accuracy lead us to other desirable properties of data. The concept which is measured is often described only very briefly (such as through the use of a term like "unemployment"). In that case, the receiver of the message may assume the concept to correspond to his or her notion of what "unemployment" is, which may or may not be the same as that which was actually and explicitly implemented. Similarly, unless an explicit statement about accuracy is provided, the receiver is free to assume any level for it, including "complete accuracy".

The result may clearly be potentially misleading. Thus potentially misleading data are data whose concepts and accuracy are inadequately or incompletely described. Misleading data are those whose concepts and accuracy are incorrectly described.

The implications of the concepts developed up to this point will be spelled out below in the context of social statistics in Canada.

3. SOCIAL SCIENCE AND SOCIAL STATISTICS

The notions outlined above, particularly the concepts of decision models, validity and relevance, help me to understand the scene of social statistics in Canada and how it developed. Let me share with you the highlights as I see them.

1. <u>Predominance of administrative data in social statistics</u>:
   A striking phenomenon of the social statistics scene in Canada is the relative predominance of data initially collected for administrative, as opposed to statistical purposes: vital statistics, statistics on health institutions, educational institutions, etc. Of course, this is not an accident. The fact is that each of these data sources (as indeed most data sources should!) came about in response to specific decision problems, typically related to the administration of particular social programs. This is not the time or place to discuss why and how the social programs themselves developed. Suffice it to say, without any value judgements being implied, that the programs themselves were largely initiated within decision models that were more political than quantitative, where the questions raised were more in the nature of "how fast can we afford to do it" rather than "what are the objectives, through what alternative means can we achieve them, and what impacts might each of the alternatives have in addition to the furtherance of the particular objective?" Thus, the launching or extension of programs,

such as universal health insurance or the large-scale expansion of higher education, did not represent decision problems[1] in support of which comprehensive statistical programs would have developed.

Having launched the programs, their efficient administration, of course, represents a continuing decision problem in need of data - the so-called administrative data. Typically, therefore, the definition of the concepts measured and the identification of the reference entities is determined within the framework of these particular decision models, e.g., the administration of hospital programs. The particular decision problems have a major impact on the data created to support them: for example, the problem of efficiently administering a set of hospitals is very different from the issue of how to improve the health of Canadians. Hospital statistics were largely developed in response to the administrative problem. While they are still relevant to a study of the general health of Canadians, their validity might clearly be impaired for some of the analyses which might be involved in such a study.

2. Problems of validity and relevance of administrative data for general purposes:

This point is implicit in the previous one. Neither the operationalized concepts used, nor the reference entities (coverage in more traditional terms) lend themselves easily to the purpose of developing or assessing general social policies, although a lot of work has been done by Statistics Canada to influence administrators to modify their concepts in such a way as to improve their validity for more general uses. A further problem of relevance of these data is the relative paucity of the data bases, i.e., the relative lack of

---

[1] In all fairness, it should be added, however, that in recent years there has been more interest and activity in the area of program evaluation, particularly before changes to major programs were introduced.

appropriate concomitant variables which would render the conceptual linkage of the data that are there with other data, and with a variety of decision models, more manageable. Record linkage, a technically difficult and often controversial undertaking, can often ameliorate this problem. On the positive side, administrative data are cheap, so long as one does not want to overcome their limitations, i.e., so long as one can use them more or less "as is". Attempts to overcome their limitations can be expensive. Given that they must be collected, any alternative collections (e.g. statistical surveys) impose not only extra costs but also response burden. And, finally, their accuracy is generally high, i.e., their accuracy as measures of their operationalized concepts, so long as one can accept the limitations of validity often resulting from those concpets.

Since the question of general relevance of data is of necessity of secondary importance to administrative agencies, they give only a passing recognition to the extent to which the resulting statistics fit decision models other than those they were initially designed to support. Nevertheless, given the relevance of administrative data sources for a broad class of users, overcoming their shortcomings for non-administrative uses is a major challenge for Statistics Canada, one to which we attach a high priority.

3. Social Surveys: If there are some decision models which are in need of statistical information not available as by-products of the administrative processes, particularly for analyses which require information cutting across the subject boundaries imposed by the institutional boundaries of administrative processes, these can only be filled by data collected specifically for statistical purposes, i.e., through statistical surveys.

As pointed out earlier, the relevance of data is partly determined by its accuracy. The requirements of accuracy, particularly if provincial or sub-provincial analyses are required, often impose a sample size which renders data collection prohibitively expensive for most non-Federal collectors. At the Federal level, the main survey vehicles used to collect social statistics are the quinquennial censuses of population and housing, the Labour Force Survey, surveys attached to it as supplementary questions whose subject matter has a largely irregular frequency, and the annual or less frequent surveys of Consumer Finance, Household Facilities and Family Expenditures. The latter three actually are either supplements to the Labour Force Survey or use its facilities some other way. During 1978-79, a new social survey, the Canada Health Survey, was launched but became a victim of budget cut-backs.

All of these surveys are widely used, at least by governments, in their decision models. A significant demonstration of the utility of the census could be seen in the reaction of data users when consideration was given to reduce the data coverage of the 1981 Census to legally mandated questions.

One can ask whether our current survey program is big, small or just about right. It is certainly small in comparison with the U.S. social statistics scene. We do not have surveys on health, comprehensive victimization, housing, time allocation, quality of life, a detailed survey to measure the impact of transfer payments, longitudinal surveys to measure the impact over time of "prevailing conditions" on cohorts which are in a state of significant transition - the pre-retirement group, or entrants into the labour force. This partial list could be extended considerably. However, I hope it is clear by now that I consider any such abstract questioning of the adequacy of our survey program to be rather futile. <u>What are the decision models which cannot be formulated; what important analyses cannot be carried out with an</u>

acceptable level of uncertainty without such data? These are the prior questions to answer. It is only by engaging in relevant analytical work that social scientists can help answer these questions.

4. Analysis of social data: In an outstanding paper delivered at the American Statistical Association, Robert Parke and Eleanor Sheldon identified several types of analysis of social data serving to render them relevant for major policy purposes. Although this section of my paper deals with highlights of the Canadian scene as I see it, nevertheless, I would like to recapitulate briefly the analytical categories they identified, using my own terminology, so as to illustrate the wide variety of important social decision models which can be assisted by suitably analyzed social data or, in fact, which can be modified by such data and analyses. The specific analytical examples are mostly American and drawn from their paper.

   (a) Cognitive information
   Studies which identify new problem areas of major social importance fall into this category. The outstanding examples are the studies linking smoking and a variety of health hazards, drinking and drug abuse, tension and auto accidents. They may not fit initially a direct programmatic decision model, but do fit the higher decision models of those responsible for health policies in general and may indirectly result in the articulation of more specific, lower level, problem-oriented decision models.

   (b) Identifying important external constraints
   The contribution here is to invite the attention of policy makers to developments not under their control that may alter the way they wish to conduct affairs that are under their control.

An example from outside the social sciences is the weather reports in which we are interested because they may alter our "adaptational strategy", even though we cannot manipulate the weather. A social science example quoted by the authors is a study of the impact of wide fluctuations in birth rates over the past 30 years on: the female labour force, low income and minority groups, geographic mobility of workers, unemployment, GNP, consumption patterns, schools, health services. Specific policies can affect each of these separate problem areas, but only within a margin determined by external constraints. The setting of realistic goals and the definition of success or failure of particular policies is an important indirect contribution.

(c) Projecting consequences

This category is somewhat similar to the previous one with the exception that manipulative, as opposed to adaptive, strategies are available. The illustration is a combination of the decline in the absolute number of births starting in 1961 with the rise in teachers' college enrollments. Good current statistics were available of both phenomena and could certainly have been suitably analyzed to forecase the inevitable over-supply of teachers. Had this been done with the appropriate penetration of the analytical results in the audiences concerned, educational authorities might well have taken appropriate actions to prevent a serious dislocation in the labour market.

(d) Analysis of specific decision options

The analysis here concentrates on specific decision options contemplated so as to assess them in terms of their contribution to the achievement of objectives. The example quoted by the authors relates to the policy objective of the early

'70's to reduce the population growth of metropolitan areas of over 1 million population by encouraging growth in smaller, less congested "growth centers". An analysis showed that between 1970 and 2000 even a 30 percent growth by such centers would only absorb about 10 million persons, leaving still 70 million persons to be absorbed by the larger metropolitan areas, mostly as a result of local births and immigration from abroad. Clearly, growth centers would not contribute significantly to the achievement of the stated objective. The decision model would need to be amended.

(e) Communicating the meaning of data

This is the well-known but insufficiently often practiced data analysis which draws out the story of the data that will not be told without it. That "story" might transform the data into information; without it the data is in danger of remaining an interpretable message. The analysis here intends to show the relevance of data for a variety of not necessarily explicitly specified decision models. Examples are: the transformation of mortality data into life expectancy tables; the presentation of income data showing the fraction of total personal income received by the one-tenth of families receiving the smallest incomes; the transformation of current marriage and divorce data into cohorts having different divorce experiences; more elaborate multivariate analyses involving standardization of populations over time with respect to some characteristics so as to study the impact of others; the use of loglinear or other techniques to study the relationship between geographic mobility and socio-demographic characteristics; etc. In case I might be misunderstood, I want to emphasize that this "story that the data tells" is <u>not</u> equivalent to the verbalization of tables, or even some analytic material that seems to be contrived and "l'art pour l'art" - the latter often strikes me as a declaration "here is the answer" without first asking "what is the question".

The reason I dwelt on the Parke-Sheldon paper at length is partly to illustrate the enormous analytical potential in social statistics, but also to make more plausible to you my belief that there is a great relative paucity of comparable analytical work in Canada. In fact, I think this is a striking aspect of the Canadian scene as compared to the U.S. One may argue that the difference is a function of the relative paucity of social statistics to analyze. With due respect, I must submit that there appears to be a shortage of analyses of even the statistics that are available. Comparing again the Canadian scene with the U.S., the federal statistical system here seems to be doing at least as good a job in disseminating its data in useable form as its U.S. counterpart. Most of our household surveys are, for example, available on tape in micro-data form, ready for analyses. The Labour Force Survey or the income surveys are examples. In the U.S., the so-called March supplements of the CPS are used by scores of users. Workshops held on them are typically over-subscribed. By contrast, as one of my colleagues (who shall remain unnamed) put it: "I wish that a few Canadian social researchers would discover the potential of the Labour Force Survey. I would estimate that an untenured assistant professor in Canada who has a logical mind and quantitative tools at his command could easily acquire tenure, promotion and a scholarly reputation by just mining the LFS".

Similar comments could be made of a number of our data bases, particularly with respect to the potential of the joint exploitation of several of our data sources.

5. <u>A climate of scarce resources and relatively low priorities</u>:
Without doubt, a major aspect of the Canadian scene of Federal social statistics is a climate of extremely scarce, in fact <u>diminishing resources</u>. It would be foolish of all of us not to recognize it and to pretend that the degrees of freedom open to us are significant.

This climate of scarce resources is general within the Federal government and certainly within Statistics Canada. The general scarcity of resources is exacerbated by the relatively lower priority accorded to social statistics compared to economic statistics. This is not a result of some mysterious internal struggle within Statistics Canada, or lack of even-handedness in the distribution of cuts. It is a direct consequence of government priorities.

Governmental priority is, of course, a complex process. It is partly established by the government's own policy agenda which usually is a judicious mix of what it believes is good for the country and what it believes the country or some important groups in it want. In the case of social statistics, government priorities can be impacted upon either by demonstrating effectively that a certain level and type of service is needed for it to be a more effective government, or that this service is vitally important for other levels of government or groups of the population which in turn are considered to be important to the government. I want to emphasize that the above should not be construed as lobbying for your vocal support. Such lobbying for more social statistics would, I believe, be ineffective at any rate. The challenge I want to put to you, and indeed to my social scientist colleagues in Statistics Canada, is for social science to make a greater and more readily visible contribution to Canadian society.

6. Social science and social statistics: The five previous highlights of the social statistics scene in Canada indicate a relatively low level of activity and even a low level of interest. The immediate question is: if there is a low level of activity and a low level of interest, perhaps supply and demand are well balanced and everything is quite alright in this best of all good worlds. Or, to quote a little poem of the Danish poet Piet Hein: "The universe may be as great as they say, but it wouldn't be missed if it didn't exist".

However, when I consider the highly relevant work quoted by Parke and Sheldon, I feel really sad about the opportunity cost to Canadian society.

I believe social science itself must become more relevant to decision makers in order for social statistics to be given higher priority. What I have in mind is the identification of social problems of recognized importance; the identification through analysis of factors related to such problems, including those which can be influenced through decisions; the determination of the extent to which the decision models cannot be articulated adequately without additional social statistics; and finally, effective communication of the fruits of social science in the language of the "man on the street".

In effect, my hypothesis is that the main issue is the image that decision makers have of social science, rather than that of social statistics. Some evidence to support this hypothesis is provided by a recent study of the Institute for Social Research, University of Michigan. The study involved "204 interviews on social science research utilization and policy formation with persons holding important positions in various departments, major agencies, and commissions of the executive branch of government". One of the results of the study is shown by Chart 2 at the end of the paper. It indicates that sociology is not rated very highly in terms of its "validity and reliability" - not only far below the "hard" sciences of physics and biology, but considerably below economics, and even below psychology. This is not due to the tools most frequently used in social statistics, as shown by Chart 3. Indeed, survey research is considered to have higher validity and reliability than controlled laboratory experiments. It does not even seem to be due to perceived limitations of social statistics. As Chart 4 shows, "population statistics" and the "unemployment rate" are rated

as very valid and reliable, more so even than economic trend data, even while economics itself is rated in Chart 2 as considerably more reliable than sociology.

The authors of the study go on to point out that "the heavier users of social science information consistently rates the social science items higher than the less frequent users and non-users. These differences in ratings across utilization score levels were sizeable and statistically significant". While one can never know, of course, the direction of causality between higher ratings and more utilization, I have sufficient confidence in social science to believe that more and more meaningful exposure by policy makers to the best that social sciences can offer will have significant benefits for both communities.

It would appear, therefore, that a prerequisite of a richer social statistics data base is greater involvement by social scientists with social problems perceived to be important by policy makers, particularly at the federal and provincial level. Put bluntly, it is not very effective for social scientists to point out to Statistics Canada that more social statistics are needed. Such demands become really effective when they come to Statistics Canada in the form of needs by public policy makers or at least for public policy articulation. An examination of the history of our major "general purpose" social statistics programs - the Census, Labour Survey, Survey of Consumer Finance, Family Expenditure Survey - will readily confirm the validity of this statement.

Let me add, in case I might be misunderstood, that there is a sharp distinction between political interference with statistics and the role of the policy process in setting statistical priorities. Political interference in statistics, however subtle, which may affect statistical data, their timeliness, or unrestricted availability is, of course, totally unacceptable. Setting statistical

priorities in response to recognized public policy needs is not only proper but the only feasible course. The provision of statistics to shed light on the priority policy issues to be tackled or monitored by our elected policy makers assists not only the government which makes the decisions, but also all others who want to participate in debates concerning those issues, or those who wish to monitor the performance of the decision makers. This is true so long as statistics, once collected, are made generally available - a condition clearly adhered to by Statistics Canada and with strong support from a long succession of governments.

4. SOME CONCLUDING REMARKS - WHERE SHOULD WE GO FROM HERE?

Much of what I said can be construed as a Statistics Canada official telling social scientists what <u>they</u> should do to improve the state of social statistics - if indeed it is not judged to be adequately developed now. This is certainly not the impression I would like to leave with you. The question is how can we, together, be of greatest assistance to those who have to grapple with the social issues facing Canadians. The following is an indication of what I see as some of the ways we can, together, be more effective.

1. First of all it bears emphasis that what I said about social scientists in general certainly applies to those of them who are employees of Statistics Canada. However, as the main source of nationally comparable social statistics, we have some additional responsibilities. So I will start this section by outlining my priorities for the next period of time as Assistant Chief Statistician of the Census and Household Surveys Field.

   (a) Statistics Canada will put a high priority on ensuring that our statistics should not be "potentially misleading". I am using this term as defined earlier, i.e. in the sense that

the concepts used and the accuracy of statistics are described and disseminated to the best of our ability. Actually, in the case of our censuses and household surveys, our record is not at all bad in this regard: in the last decade such information has consistently been produced and made available. This practice will continue, but we will put more effort on calling attention to the availability of such information as opposed to simply making it available. I have the feeling that a good deal of material we produce on the accuracy of our data or their validity for different purposes would qualify as "interpretable messages", duly put into the public domain but becoming "information" to only a few users.

(b) I will attempt to strike a new balance between resources spent on generating data and those devoted to transforming them into issue-oriented social information. This applies not only to the so-called general purpose statistics, but also to statistics derived from administrative records. As indicated earlier, many of the latter files have high potential relevance in relation to a number of issues, even though their concepts and coverage diminish their validity for some purposes. Our analytical work should not consist of bland verbalizations of tables or purposeless researching of relationships which are not consciously designed to shed light on important hypotheses, issues or data problems. At the time time our analytical work must be carefully balanced and monitored to ensure that it remain objective, i.e. that policy relevance is maintained without policy advocacy. The distinction is subtle but, for a government statistical agency, extremely important: for example, we could analyze the extent to which unemployment, as measured, represents a hardship for different subgroups of the population - not the impact of alternative unemployment insurance schemes. To preserve

their objectivity, similar rules will apply to data analysis as to statistical data: they will be publicly available without preference to any group of users, and their assumptions, methods, and data sources will be clearly stated.

(c) In order to facilitate analytical work to be carried out by the academic community and others outside of Statistics Canada, I will bend every effort to make our data available and accessible. Subject to the very real resource constraint described earlier, I will stress the building of bridges with our user communities, particularly those who will help to render our data more meaningful and relevant to the public. Much of our data is available on micro-data tapes; all of it (subject to confidentiality constraints) is available for special retrievals. Perhaps, we can do more to demonstrate the potential of such data bases or otherwise assist analysts outside of Statistics Canada. I have a number of concrete ideas in mind and am also more than willing to listen to suggestions.

2. It is presumptuous of me to tell you what you should do. Nevertheless, being concerned with the status of social science and social statistics in Canada, I know what I would do if I were a social scientist. I would identify for myself a set of socially important issues, formulate some hypotheses which might shed some new light on them, perhaps point the way to some alternative remedies or show the unworkability of others, in other words engage in one of the types of policy-relevant analytical work outlined in the Parke-Sheldon paper.

3. It might be argued that the notion of more, and more relevant, analytical work as a prerequisite for a richer social data base may appear to be somewhat "upside down": if important data sources are missing, how can we do relevant analyses? I believe we have a long way to go before we can say that the existing data base is reasonably well exploited in terms of its analytical potential. Furthermore, it is only in the context of developing relevant decision models that the importance of data gaps can be convincingly demonstrated: by showing how a competent, relevant analysis could have been even better if specific statistics had been available. In the current tight economic climate, I do not want to give you the impression that even if a good case is made for certain types of statistics, the funds will be made available for them. However, without such a case the funds will certainly not be made available. Furthermore, I foresee a good deal of gradualism. When the need for new statistics is convincingly demonstrated, including the social issues they are supposed to help address, we will have to confirm (or otherwise) such findings through intensively exploited small scale surveys. Let such small scale surveys begin to be useful and relevant and let the limitation of sample size and/or frequency of measurement be seen to be a deterrent to more effective decision making before such surveys are expanded.

4. I was struck in the United States by the close ties, both formal and informal, which exist between the Federal statistical system and the academic community. It is a healthy relationship, not of the "you scratch my back, I'll scratch yours" variety. As the case may be, the academic community is either a vociferous critic or a most influential supporter of the statistical system. One thing it is not: a neutral bystander. My early impression is that, with a few notable exceptions, the academic community is far too passive in Canada in relation to the Federal statistical system. We need to have strong, fair, informed and influential critics. Without them the pressure to improve the quality of our

product is largely internally generated. Other users put pressures on us to improve the timeliness and quantity of statistics produced, and to improve our methods of dissemination. It is my impression that not enough external pressure is put on us to improve the accuracy of our data. Even the most professional organization is strongly tempted to respond more readily to external rather than internal pressures. I should add that the most useful critics are those who understand thoroughly our statistics, "warts and all", as well as the constraints under which we operate.

We also need to have your strong support when the system as a whole is in danger - for whatever reason. It must be kept in mind that the value of information accrues only through its use; but the cost of it is entirely front loaded. Moreover, information cannot be divided up so that each unit of it could be marketed at a relatively small cost: the entire cost must be prepaid. Under these circumstances, the "buyer" of information, basically, must rely on the reputation of the producer. Other facts are also relevant: large-scale statistical data collection, because of its cost, tends to be a government monopoly; we are asking the public to give us of their time and to share with us information that many consider to be confidential; and we are asking those participating in democratic debates to accept statistical information as a factual base for such debates. All of these factors together put a high premium on preserving the reputation of, and confidence in, the statistical system. I submit to you that when that is put into question, it is not in your own interest to be idle bystanders. To borrow a Latin phrase, "nostra res agitur" or, loosely translated, we are all involved together. If there is a real problem, to into battle to isolate and remove it; if there isn't, let it be known loud and clear. Should the public confidence in the statistical system ever be shaken, then to talk about future developments of social statistics amounts to fiddling while Rome burns.

5. Finally, I want to sound a note of caution: in some sense we can be too successful in rendering statistics policy relevant. I have in mind the disconcerting trend, particularly in the U.S. but to some extent here as well, of using statistics for formula-based decisions. Throughout this paper, at several points, I emphasized that statistical data become information through interpretation. I made several references to decision models. Without interpretation, statistics become just data; without judgement, decision models become formulae. Yet in a society in which decision making is increasingly subject to public scrutiny - by itself a largely healthy development - but which also is increasingly reluctant to accept judgements as the bases of public decisions, there is something very attractive in formula-based decisions. The tendency to try to render decisions unchallengeable is not restricted to the use of statistics. To give just two other examples, one finds instances of it in the unthinking use of high school marks as the basis of university admissions, or in the tendency of some medical doctors to obtain "irrefutable" evidence for their diagnoses by prescribing more laboratory tests than might be necessary without the increasing risk of malpractice suits. What concerns me here, however, is the use of statistics in formula decisions for purposes they were not designed to support. In the U.S., such uses often resulted in some extra funding support for new or expanded statistical data collections. However, I believe that the resulting politicization of statistics is too high a price to pay for the benefits involved.

In closing, I hope you will forgive my frankness. However, I believe that so shortly after having accepted my new position, I have a unique opportunity: to the extent you agree with some of my points, I am off to a good start; to the extent I might get into hot water with you, I can always claim ignorance. This opportunity may never return.

RESUME

Le présent document examine la situation actuelle de la statistique sociale au Canada. Il définit et analyse certains concepts fondamentaux relatifs à la statistique et à l'information sociale. L'accent est mis sur la disponibilité et l'analyse des données administratives, ainsi que sur la nécessité des enquêtes sociales. L'auteur propose différentes analyses de données susceptibles de permettre l'élaboration de modèles de décision sociale permettant l'application des politiques. Enfin, le document décrit dans les grandes lignes le travail à accomplir en priorité en vue d'une meilleure utilisation de la statistique sociale.

CHART 1

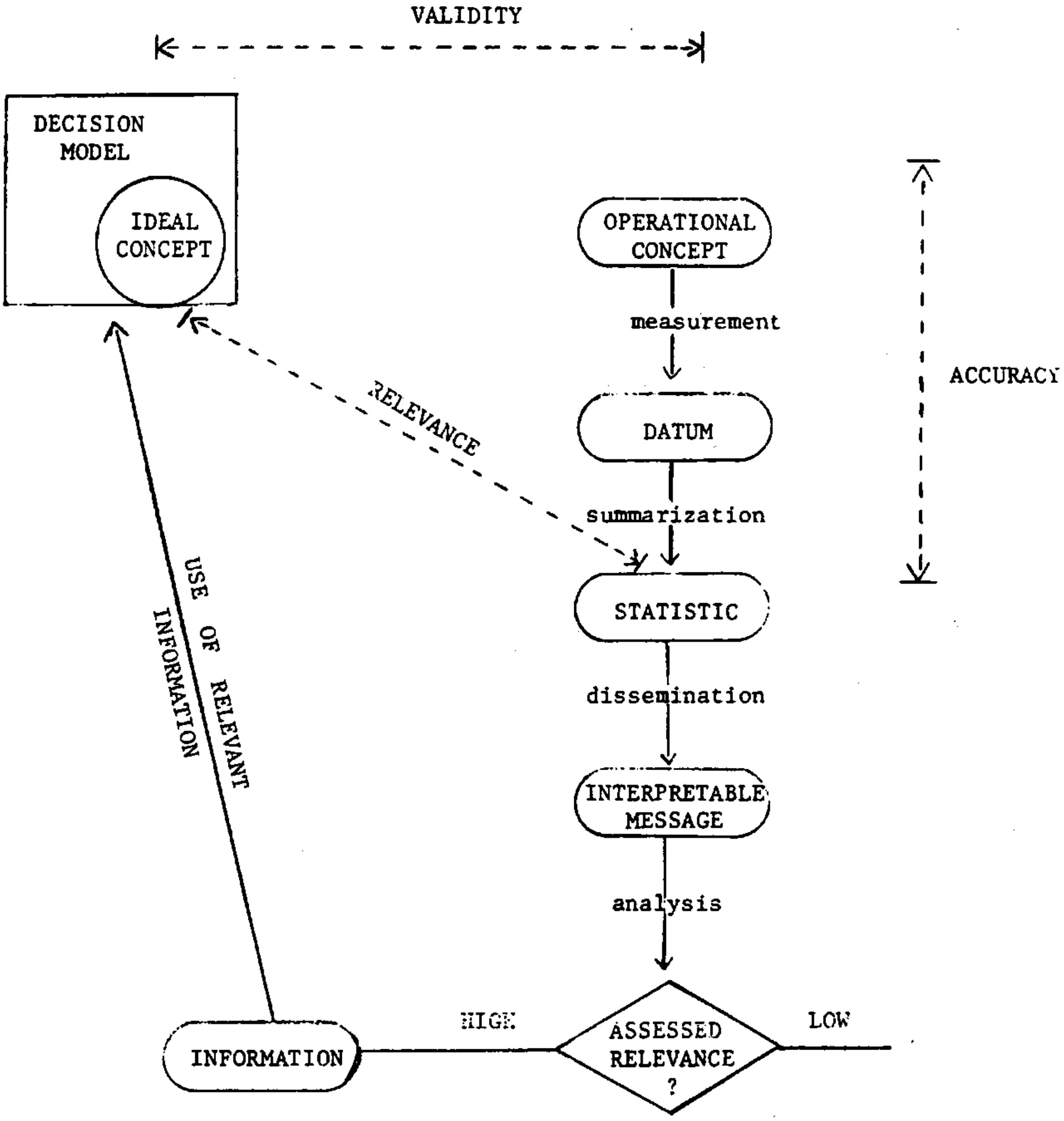
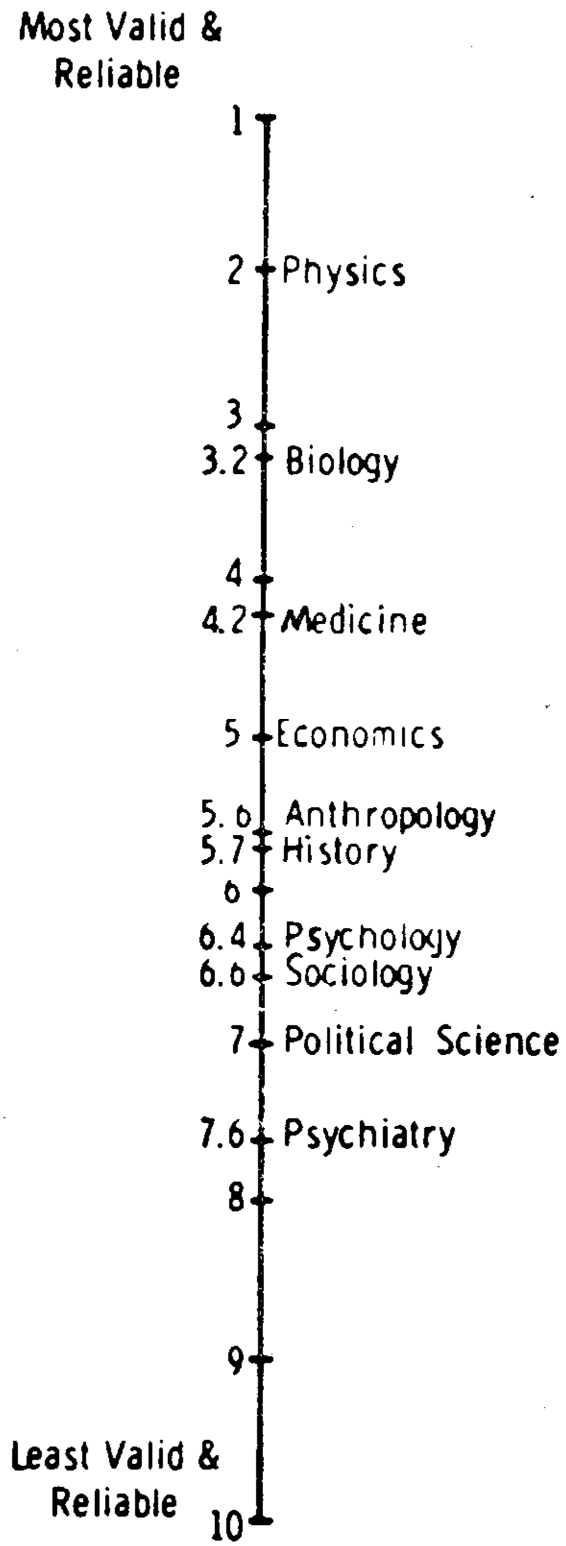
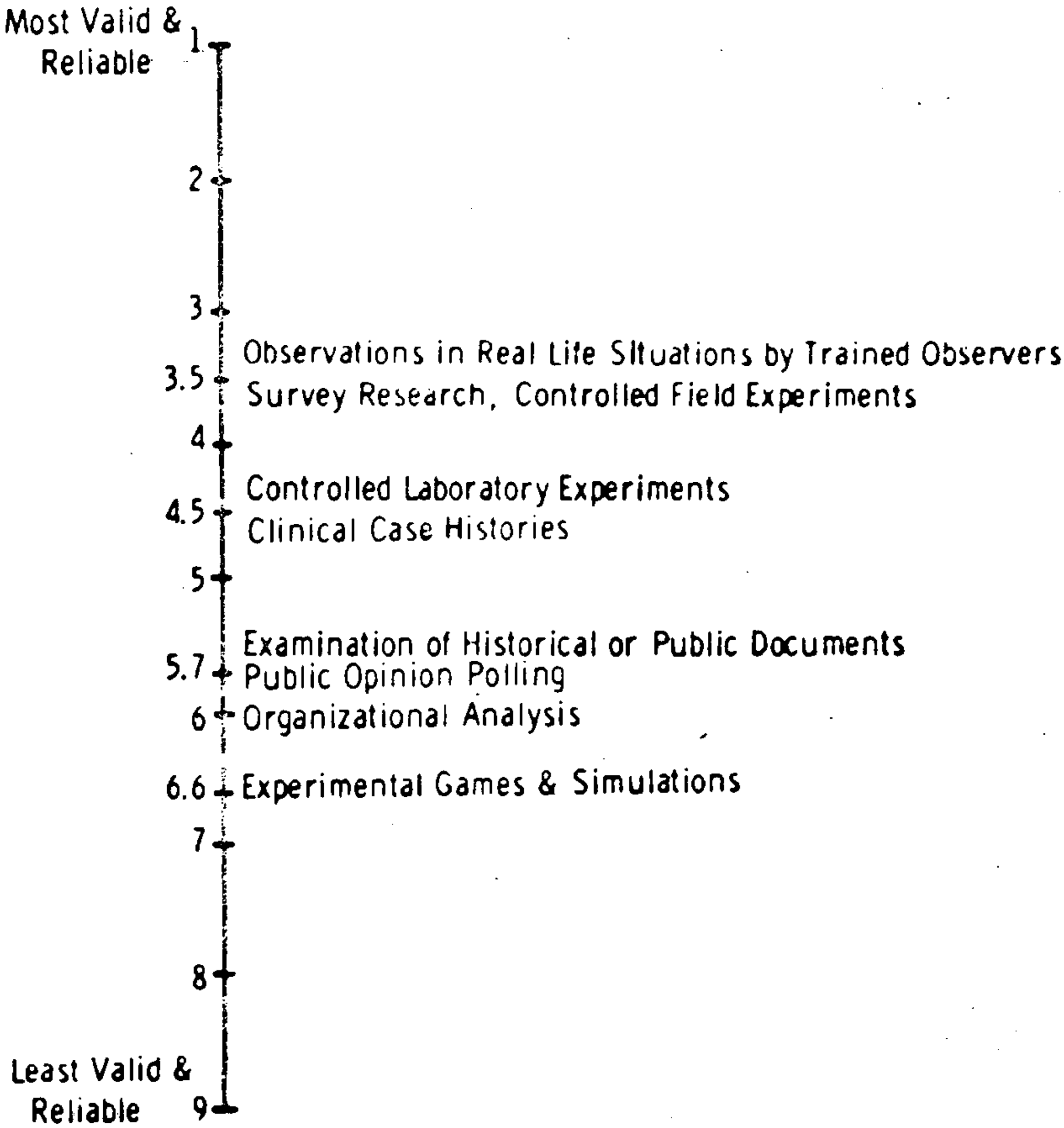

VALIDITY
DECISION MODEL
IDEAL CONCEPT
OPERATIONAL CONCEPT
measurement
DATUM
summarization
STATISTIC
dissemination
INTERPRETABLE MESSAGE
analysis
ASSESSED RELEVANCE ?
HIGH
LOW
INFORMATION
USE OF RELEVANT INFORMATION
RELEVANCE
ACCURACY

Chart 2



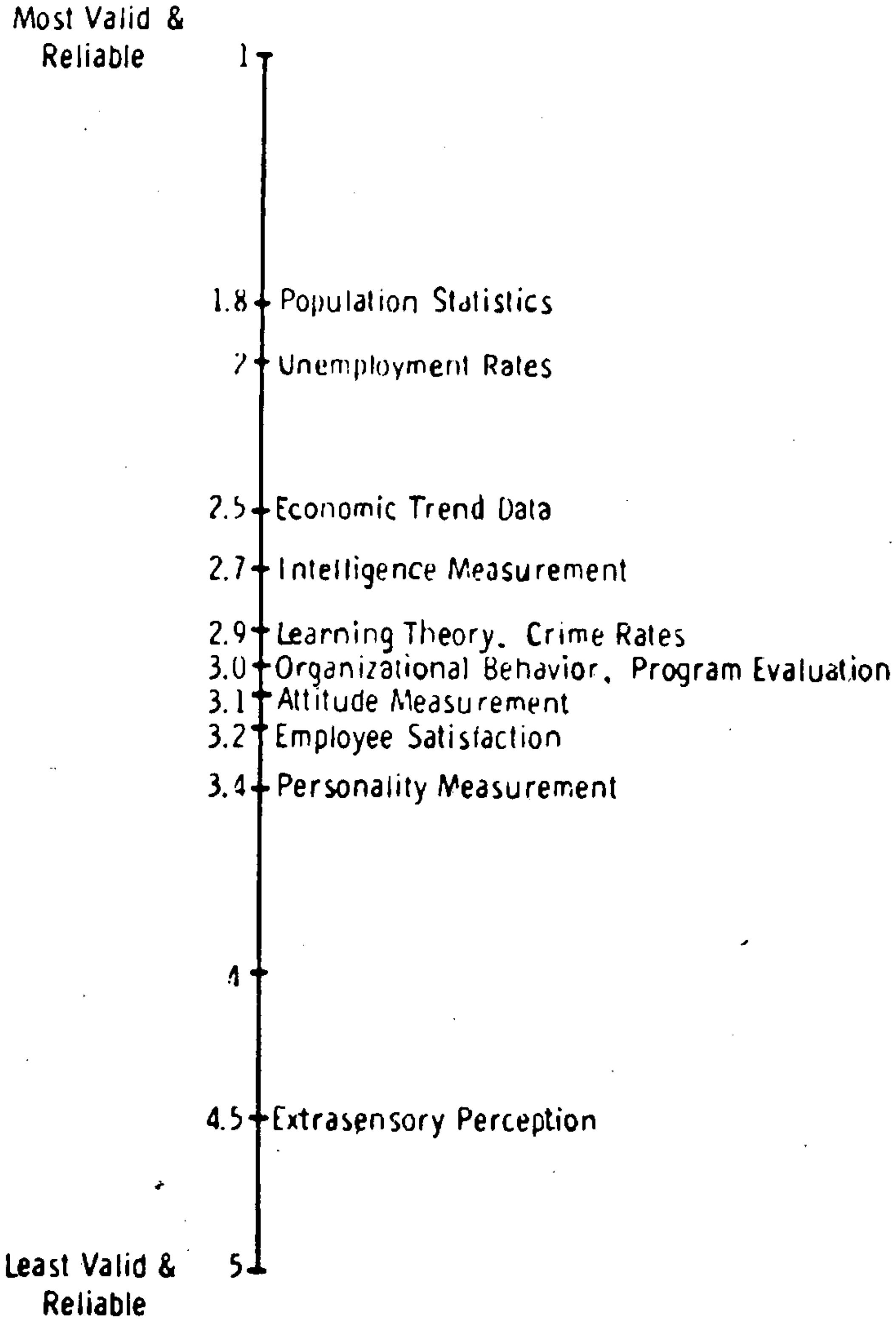
Mean Ranking of Validity and Reliability for Scientific Disciplines

Chart 3



Mean Rankings of Various Data Gathering Techniques

Chart 4



Mean Validity & Reliability Ratings for Various Types of Data

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE 1979, VOL. 5 NO. 2

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT - A REJECTIVE METHOD

G.H. Choudhry and M.P. Singh[1]

An alternative to the direct selection of sample is suggested, which while retaining the efficiency at the same level simplifies the selection and variance estimation processes in a wide variety of situations. If n* is the largest feasible ΠPS sample size that can be drawn from a given population of size N, then the proposed method entails selection of m (=N-n*) units using a ΠPS scheme and rejecting these units from the population such that the remainder is a ΠPS sample of n* units; the final sample of n units is then selected as a subsample from the remainder set. This method for selecting the ΠPS sample can be seen as an analogue of SRS where it is well known that the "unsampled" part of the population as well as any subsample from this part are also SRS from the entire population when SRS is the procedure used. The method is very practical for situations where m is less than the actual sample size n. Moreover, the method has the additional advantage in the context of continuing surveys, eg. Canadian Labour Force Survey (LFS), where the number of primary sampling units (PSU's) may have to be increased (or decreased) subsequent to the initial selection of the sample. The method also has advantages in the case of sample rotation. Main features of the proposed scheme and its limitations are given. Efficiency of the method is also evaluated empirically.

1. INTRODUCTION

Selection of primary sampling units (PSU's in a multi-stage sampling scheme) with unequal probabilities has found wide applications in large scale surveys. However, in many cases either with replacement or one PSU per stratum sampling is used because of their simplicity. The more efficient without replacement sampling schemes (e.g. Fellegi [1963],

---

[1] G.H. Choudhry and M.P. Singh, Census and Household Survey Methods Division, Statistics Canada.

Hartley and Rao [1962], etc.) become quite complex either in terms of the selection procedure or variance estimation, even for moderately large sample sizes. In this article we propose an alternative approach (a rejective method) of selecting the sample which retains the efficiency of direct selection method using any sampling scheme for the purpose of rejecting m units. This method simplifies the sample selection and usually the variance estimation in a wide variety of situations. In addition, the suggested approach has several other operational advantages in the context of large scale continuing surveys such as:

a) changes in the sample, both in terms of increase or decrease in the number of PSU's depending upon the need of the time, can be achieved just as easily as in the case of simple random sampling (SRS), and

b) large scale survey frames are often used as main sources of selecting samples for ad hoc surveys from time to time (see Drew, Choudhry and Gray [1978]).

In such cases, following this approach, unequal probability samples can be selected for the ad hoc surveys with the same ease as SRS and without conflicting with the main continuous survey, once the samples has been selected for the main survey.

c) A feature often required for continuous surveys is the rotation of PSU's after a certain period of time. Again, this can easily be achieved in this approach irrespective of the complexity of the selection scheme used (for rejecting m units).

This rejective method used for selecting ΠPS samples is very practical and is recommended especially for those situations where the value of the parameter m is less than the sample size n, since the computation of Π_{ij}, the probability that both the units i and j are in the sample, is

simplified to a great extent. It should be emphasized that if m is greater than the sample size n, then one should use the direct ΠPS method for selecting the sample unless there are over-riding considerations such as those mentioned earlier in the context of continuing surveys.

Sections 2 and 3 describe the actual selection mechanism and calculations of the inclusion probabilities and the joint inclusion probabilities. In section 4, suitability of this approach for selecting PSU's in Non Self-Representing (NSR) areas in the Canadian Labour Force Survey (LFS) is demonstrated. Results of empirical study are presented in the last section comparing the efficiencies of the Horvitz-Thompson [1952] estimator.

2. SELECTION PROCEDURE

The given finite population U consists of N units, $\{u_1, u_2, \ldots, u_N\}$ and a known "size measure" x_i is associated with the unit u_i; i=1, 2, ..., N. It is required to draw a sample of n distinct units from the population in such a way that the probability of unit u_i being in the sample is proportional to its size x_i (ΠPS) for each i=1, 2, ..., N.

Define the "Normalized Sizes" p_i; i=1, 2, ..., N such that $\sum_{i=1}^{N} p_i = 1$, i.e.

$$p_i = \frac{x_i}{\sum_{i=1}^{N} x_i} \quad ; \; i=1, 2, \ldots N. \qquad (2.1)$$

A sample of n distinct units will be selected from the population such that the probability Π_i; (i=1, 2, ..., N) for the ith unit to be in the sample is np_i. Since Π_i; i=1, 2, ..., N are the probabilities and

hence necessarily less than or equal to one, therefore, the largest possible value of n (say n*) is given by

$$n^* = \left[\frac{1}{p_{(N)}}\right], \tag{2.2}$$

where $p_{(N)}$ = Max $(p_1, p_2, \ldots, p_N)$ and $[\cdot]$ is the integer function, i.e. the function gives the largest integer less than or equal to the argument.

The first step in the proposed method is to select $m(=N-n^*)$ units from the population using any ΠPS scheme and reject these units such that the remainder is a ΠPS sample of size n^* from the given population. A simple random sample without replacement (SRSWOR) of n out of n^* retained units is then selected. In order to show that this final sample is a ΠPS sample of size n from the entire population, we define

$$p_i^* = \frac{1-n^* p_i}{N-n^*}\,; \quad i=1,2, \ldots, N\,. \tag{2.3}$$

It can be readily checked that p_i^*, $(i=1,2, \ldots, N)$ are the probabilities since from (2.3) $p_i^* > 0$ for all i and also from (2.3) $\sum_{i=1}^{N} p_i^* = 1$, therefore, $0 < p_i^* < 1$, $i=1, 2, \ldots, N$.

Since the sample of size m is selected using any given ΠPS scheme, the probability that the unit i is in the sample is mp_i^*; that is:

$$\Pr(i \varepsilon R) = mp_i^*\,; \qquad i=1, 2, \ldots, N, \tag{2.4}$$

where $R = \{R_1, R_2, \ldots, R_m\}$ is the set of m units selected with sizes p_i^*, $i=1, 2, \ldots, N$ using a ΠPS scheme. Let S^* be the set of n^* units not in the R, i.e. $S^*=U-R$. Now the final sample is a simple random sample of n units out of n^* units in S^*. We denote by S the set of n units

selected in the second stage of sampling, i.e. SRS of n out of n^* units in S^*. That S is a ΠPS sample of size n from the population U of size N can be shown as follows:

$$Pr(i \varepsilon S^*) = Pr(i \notin R)$$

$$= 1 - mp_i^* ; \qquad i=1, 2, \ldots, N. \qquad (2.5)$$

Substituting $m=N-n^*$ and p_i^* from (2.2) in equation (2.5) gives

$$Pr(i \varepsilon S^*) = n^* p_i ; \qquad i=1, 2, \ldots, N. \qquad (2.6)$$

Thus, S^* is a ΠPS sample of size n^* (largest permissible sample size) from U. Since S is SRS of n out of n^* in S^*, therefore, we have

$$Pr(i \varepsilon S) = Pr(i \varepsilon S^*) \times Pr(i \varepsilon S | i \varepsilon S^*)$$

$$= (n^* p_i) \times (\frac{n}{n^*})$$

$$= np_i \quad ; \qquad i=1, 2, \ldots, N. \qquad (2.7)$$

Denoting by Π_i, the probability that the unit i is in the set S we write

$$\Pi_i = np_i \quad ; \qquad i=1, 2, \ldots, N, \qquad (2.8)$$

i.e. S is a ΠPS sample of size n from the population U of size N.

When m<<n, the sample of m units can be selected using, for example, Fellegi's method [1963] or randomized PPS systematic method due to Hartley and Rao [1962] depending on the value of the parameter m and the population size N. Sinha [1973] has also suggested a rejective sampling scheme which achieves pre-specified inclusion probabilities of first two orders. Suitability of other methods may also be investigated for various situations.

3. CALCULATION OF Π_{ij}

The formula for Π_{ij}, the probability that both the units i and j are in the sample denoted by S in the previous section, will be derived. Denoting by Π^*_{ij}, the probability that both the units i and j are in S^*, we immediately have:

$$\Pi_{ij} = \frac{n(n-1)}{n^*(n^*-1)} \times \Pi^*_{ij}; \quad i=1,\ 2,\ \ldots,\ N-1, \quad j=i+1,\ i+2,\ \ldots\ N. \tag{3.1}$$

In order to find Π^*_{ij}, we define the following four mutually exclusive and exhaustive events for the units i and j with respect to the sets R and S^*.

Event E1: Both the units i and j in R.
Event E2: Unit i in R and unit j in S^*.
Event E3: Unit i in S^* and unit j in R.
Event E4: Both the units i and j in S^*.

Then we have:

$$\Pi^*_{ij} = Pr(E4)$$

$$= 1-\{Pr(E1) + Pr(E2) + Pr(E3)\}\ . \tag{3.2}$$

But $Pr(E1) + Pr(E2) = Pr(i\varepsilon R)$

$$= mp^*_i\ . \tag{3.3}$$

Similarly

$$Pr(E1) + Pr(E3) = Pr(j\varepsilon R)$$

$$= mp^*_j\ . \tag{3.4}$$

Adding (3.3) and (3.4) gives

$$\Pr(E1) + \Pr(E1) + \Pr(E2) + \Pr(E3) = m(p_i^* + p_j^*). \qquad (3.5)$$

Substituting in (3.2) for $\Pr(E1) + \Pr(E2) + \Pr(E3)$ from (3.5) we obtain

$$\Pi_{ij}^* = 1 - m(p_i^* + p_i^*) + \Pr(E1)$$

$$= 1 - m(p_i^* + p_i^*) + \delta_{ij} \qquad (3.6)$$

where

$$\delta_{ij} = \Pr(E1)$$

$$= \Pr(\text{Both } i,\ j \varepsilon R).$$

Substituting for Π_{ij}^* from (3.6) above in (3.1) gives Π_{ij}, the probability that both i, j are in the set S, i.e.

$$\Pi_{ij} = \frac{n(n-1)}{n^*(n^*-1)} \times [1 - m(p_i^* + p_j^*) + \delta_{ij}]; \qquad (3.7)$$

$$i = 1,\ 2,\ \ldots,\ N-1,$$
$$j = i+1,\ i+2,\ \ldots,\ N,$$

where $m = N - n^*$ and p_i^*, $(i = 1, 2, \ldots, N)$ is defined in (2.3).

In conventional (or direct) sampling, as n increases, the computation of Π_{ij} becomes complicated, but in this case, the complexity lies only in the computation of δ_{ij}, which depends on the value of the parameter m. Since m is a population parameter and does not depend on the size of the sample, the complexity of Π_{ij} will not increase with the size of the sample. It may be noted that for the special case when the value of the

parameter m is equal to 1, $\delta_{ij} = 0$, thus

$$\Pi_{ij}^* = 1-(p_i^*+p_i^*)$$

$$= (N-1)(p_i+p_j)-1 \tag{3.8}$$

and

$$\Pi_{ij} = \frac{n(n-1)}{(N-1)(N-2)} [(N-1)(p_i+p_j)-1]. \tag{3.9}$$

Since in this case only one unit (m=1) will be rejected with PPS, therefore, no special sampling scheme is required with this rejective method.

4. APPLICATION TO CANADIAN LABOUR FORCE SURVEY

The Canadian Labour Force Survey (LFS) follows a stratified multi-stage sampling design [see Statistics Canada, Catalogue No. 71-526]. In the non self-representing (NSR) areas, comprised of rural areas and small urban centers, a ΠPS sample of PSU's is selected from each stratum, where the "size measure" is the total population of the PSU from the previous census. In the earlier design, prior to the 1971 redesign, Fellegi's [1963] method was used to select PSU's where two PSU's were selected from each stratum. The method, though quite efficient, becomes very complicated when the number of PSU's to be selected is large, usually more than three. During the redesign of the LFS following the 1971 Census, one of the criterion for the choice of selection procedure at the design stage was that the procedure should be flexible enough to allow expansion of the sample in terms of the number of PSU's as well as rotation of PSU's. The randomized PPS (probability proportional to size) systematic sampling method [Hartley and Rao, 1962] was adopted for selecting PSU's, since it meets the necessary requirements [Gray, 1973]. However, computation of Π_{ij} becomes complicated for large values of n and N. This proposed sampling scheme,

while equally efficient as well, has the additional advantage of simplicity for sample expansion and computation of Π_{ij}'s.

Following the redesign of the LFS, the sample was expanded in terms of number of PSU's in the NSR areas and the current number of selected PSU's per stratum varies from province to province. The actual number of PSU's selected from each stratum within a province and the number of cases with m<n by province are given in Table A1 (Appendix A). In the context of simplifying the calculations of the joint inclusion probabilities for variance estimation, one would see that out of 127 NSR strata across Canada, 108 would result in simpler calculations, 16 with equal difficulty and only 3 with greater difficulty. Since the proposed scheme has the advantage that it permits the sample increase by simply selecting additional PSU's with SRS, thus the procedure among others may be considered for selecting PSU's in NSR areas during the redesign of LFS following the 1981 Census. Moreover, after the sample increase, the new Π_{ij}'s will be reconstructed from the previous ones by simply multiplying with the appropriate constant factor.

5. EMPIRICAL STUDY

We have chosen 4 populations to empirically evaluate the rejective scheme of sampling for samples of size 2, 3 and 4. The description of these populations is given in Table 1, where x is the known value of the size measure, and y is the value of the characteristic of interest which is unknown but measurable. The value of the parameter m is also given in Table 1 for each of the populations.

Table 1: Description of the Populations for Empirical Study

Pop.No.	Source	N	m	CV(x)[1]	CV(y)	ρ_{xy}
1	Fellegi [1963]	6	2	0.25	0.64	0.93
2	Gray [1971 b]	10	2	0.08	0.07	0.93
3	Cochran [1963, P. 204]	10	3	0.17	0.19	0.97
4	Cochran [1963, P. 225]	10	2	0.14	0.15	0.65

1

$$CV(x) = \frac{\sqrt{\sum_{i=1}^{N} (x_i-\bar{x})^2/N}}{\bar{x}}$$

CV(y) is defined in the same fashion.

In this study, the rejective scheme of sampling is accomplished by rejecting m units with Fellegi's [1963] PPS method, and also by rejecting m units with Randomized PPS systematic method of Hartley and Rao [1962]. The Horvitz and Thompson [1952] estimator $\hat{Y}_{HT} = \frac{1}{n} \sum_{i \varepsilon S} y_i/p_i$, where $\sum_{i \varepsilon S}$ denotes the sum over the n units in the sample, for estimating the unknown population total $Y = \sum_{i=1}^{N} y_i$ is considered. The variance of $\hat{Y}_{HT}$ as given by Yates and Grundy [1953], i.e.

$$V(\hat{Y}_{HT}) = \frac{1}{n^2} \sum_{i<j}\sum (\Pi_i\Pi_j-\Pi_{ij})(\frac{y_i}{p_i} - \frac{y_j}{p_j})^2 \qquad (5.1)$$

was computed for the rejective scheme when m units are rejected with Fellegi's PPS method and also when m units are rejected with Randomized PPS Systematic method, and these variances were then compared with those when sampling directly with Fellegi's PPS method and the Randomized PPS

Systematic method respectively. In order to compute the joint probabilities Π_{ij}'s for Fellegi's PPS method, the "working probabilities" of the method were computed by an iterative procedure and the Π_{ij}'s were constructed by summing the probabilities of all those samples that contain both the units i and j. For sample sizes greater than 4, the procedure becomes more complicated and involves a great deal of tedious calculations. For the Randomized PPS Systematic method, the Π_{ij}'s were computed using a FORTRAN subroutine by Hidiroglou and Gray [1979]. The algorithm used by the above authors is a modification of Connor's [1966] formula and is due to Gray [1971a]. Variances for the rejective method using Fellegi's method to reject m units and for Fellegi's method for selecting the sample for samples of size 2, 3, and 4 are given in Table B1 (Appendix B). Similar comparison is made by replacing Fellegi's method by Randomized PPS Systematic method both for rejecting m units in the rejective method and for selecting the sample, and these results are given in Table B2 (Appendix B).

From the two tables in Appendix B, it is seen that the rejective method has the same level of efficiency as the PPS method used in the rejective method to reject the m units, i.e. Fellegi's PPS method and the Randomized PPS Systematic method. Moreover, a comparison between Tables B1 and B2 shows that Fellegi's PPS method and the Randomized PPS Systematic method have the same variances and therefore are equally efficient. Since the efficiency of the rejective method is the same as the PPS method used in the rejective method to reject m units, there will be an advantage in using the rejective method for those populations where the value of the parameter m is less than or equal to the actual sample size n.

ACKNOWLEDGMENT

The authors wish to thank the referee for some useful comments.

RESUME

On propose en remplacement de la sélection directe de l'échantillon une autre solution qui, tout en maintenant l'efficacité au même niveau, simplifie les processus de sélection et d'estimation des variances dans un grand nombre de cas. Si n* représente la plus grande taille possible de l'échantillon prélevé selon une méthode qui donne à chaque unité une probabilité d'inclusion proportionnelle à sa taille (ΠPT) à partir d'une population donnée de taille N, la méthode proposée suppose alors la sélection des unités m (=N-n*) en utilisant le schéma ΠPT et en retirant ces unités de la population de manière à ce que le reste soit un échantillon ΠPT d'unités n*; l'échantillon définitif des unités n est ensuite prélevé comme sous-échantillon à partir de l'ensemble restant. Cette méthode de sélection de l'échantillon ΠPT peut être considérée comme l'équivalent de l'EAS dans lequel il est bien connu que la partie 'non échantillonnée' de la population et tout sous-échantillon de cette partie constituent également l'EAS de l'ensemble de la population, si l'on applique la procédure EAS. La méthode est très pratique dans les cas où m est inférieur à la taille réelle n de l'échantillon. De plus, elle présente un autre avantage pour les enquêtes permanentes, par exemple l'enquête sur la population active du Canada (EPA) où il faut augmenter (ou diminuer) le nombre des unités primaires d'échantillonnage (UPE) après la sélection initiale de l'échantillon. La méthode est également intéressante dans le cas du renouvellement de l'échantillon. Le document présente les avantages et inconvénients du plan proposé. L'efficacité de la méthode y est aussi évaluée de façon empirique.

REFERENCES

[1] Cochran, W.G., (1963), Sampling Techniques, Second Edition, New York: John Wiley & Sons.

[2] Connor, W.S. (1966), "An Exact Formula for the Probability that Two Specified Sampling Units Will Occur in a Sample Drawn with Unequal Probabilities and Without Replacement", Journal of the American Statistical Association, 61, 384-390.

[3] Drew, J.D., Choudhry, G.H. and Gray, G.B. (1978), "Some Methods For Updating Sample Survey Frames and Their Effects on Estimation", Survey Methodology, Statistics Canada, 4, 225-263.

[4] Fellegi, I.P. (1963), "Sampling With Varying Probabilities Without Replacement: Rotating and Non-Rotating Samples", Journal of the American Statistical Association, 58, 183-201.

[5] Gray, G.B. (1971), "Joint Probabilities of Selection of Units in Systematic Samples", Proceedings of the American Statistical.

[6] Gray, G.B. (1971), "Variance Components and Variance Functions", Proceedings of 'Statistics Canada' 71, 119-126.

[7] Gray, G.B. (1973), "On Increasing the Sample Size (number of PSU's)", Internal Technical Memorandum, Census and Household Survey Methods Division, Statistics Canada.

[8] Hartley, H.O. and Rao, J.N.K. (1962), "Sampling With Unequal Probabilities and Without Replacement", Annals of Mathematical Statistics, 33, 350-374.

[9] Hidiroglou, M.A. and Gray, G.B. (1979), "Construction of Joint Probability of Seleciton for Systematic PPS Sampling", Journal of the Royal Statistical Society.

[10] Horvitz, D.G. and Thompson, D.J. (1952), "A Generalization Of Sampling Without Replacement from a Finite Universe", Journal of the American Statistical Association, 47, 663-685.

[11] Sinha, B.K. (1973), "On Sampling Schemes to Realize Pre-assigned Sets of Inclusion Probabilities of First Two Orders", Calcutta Statistical Association Bulletin, 22, 89-110.

[12] Statistics Canada (1977), "Methodology of the Canadian Labour Force Survey (1976)", Catalogue No. 71-526.

[13] Yates, F. and Grundy, P.M. (1953), "Selection Without Replacement from Within Strata With Probability Proportional to Size", Journal of the Royal Statistical Society, Series B, 15, 235-261.

Appendix A

Table A1: Number of Strata With Rejective Sample of Units Compared With Actual Number Selected From NSR Strata in The Canadian Labour Force Survey

		No. of Strata with m<, = or > n			No. of Strata with m equal to n			
Province	n	m<n	m=n	m>n	1	2	3	4
Newfoundland	4	11	1	0	6	5	2	1
Prince Edward Island	6	3	0	0	0	3	0	0
Nova Scotia	3	9	5	1	3	6	5	1
New Brunswick	3	10	4	0	2	8	4	0
Quebec	3	18	1	2	3	15	1	2
Ontario	3	15	5	0	3	12	5	0
Manitoba	6	7	0	0	4	3	0	0
Saskatchewan	6	11	0	0	2	5	4	0
Alberta	4	14	0	0	9	5	0	0
British Columbia	4	10	0	0	4	5	1	0
Canada		108 less[1] difficulty	16 equal difficulty	3 greater difficulty	34	67	22	4

[1] Less, equal or greater difficulty refers to calculations of joint inclusion probabilities.

Appendix B

Table B1: Variances for the Rejective Scheme Using Fellegi's Method For Rejecting m Units (Scheme 1) and for Fellegi's Method For Selecting the Sample (Scheme 2)

Pop.No.	Sampling Schene	Sample Size 2	3	4
1	Scheme 1	8.0161	3.6782	1.5092
	Scheme 2	8.1672	3.8258	1.5269
	Efficiency (1 vs 2)	101.88%	104.01%	101.17%
2	Scheme 1	3.4922	2.0475	1.3251
	Scheme 2	3.4948	2.0509	1.3287
	Efficiency (1 vs 2)	100.07%	100.17%	100.27%
3	Scheme 1	276.04	161.67	104.48
	Scheme 2	276.15	161.81	104.63
	Efficiency (1 vs 2)	100.04%	100.09%	100.14%
4	Scheme 1	6375.5	3756.6	2447.1
	Scheme 2	6373.2	3753.7	2444.1
	Efficiency (1 vs 2)	99.96%	99.92%	99.88%

Table B2: Variances for the Rejective Scheme Using Randomized PPS Systematic Method for Rejecting m Units (Scheme 3) and for Randomized PPS Systematic Method for Selecting the Sample (Scheme 4)

Pop.No.	Sampling Scheme	Sample Size 2	3	4
1	Scheme 3	8.0261	3.6915	1.5242
	Scheme 4	8.5073	4.3927	1.5242
	Efficiency (3 vs 4)	106.00%	119.00%	100.00%
2	Scheme 3	3.4922	2.0475	1.3251
	Scheme 4	3.5114	2.0423	1.3309
	Efficiency (3 vs 4)	100.55%	99.75%	100.44%
3	Scheme 3	276.41	162.16	105.03
	Scheme 4	276.20	160.23	103.65
	Efficiency (3 vs 4)	99.92%	98.81%	98.69%
4	Scheme 3	6376.6	3758.0	2448.7
	Scheme 4	6373.7	3750.7	2446.2
	Efficiency (3 vs 4)	99.95%	99.81%	99.90%

TEST OF MULTIPLE FRAME SAMPLING TECHNIQUES FOR AGRICULTURAL SURVEYS: NEW BRUNSWICK, 1978

B. Armstrong[1]

The problem considered in this paper is the estimation of various agricultural variables using a multiple frame approach. The list frame is completely contained within the area frame. The stratification for the list and area frames are based on different criteria. Overall, the multiple frame shows some gains in terms of variance over the area frame. However, a more careful analysis reveals problem areas associated with the list frame such as the method of stratification and the degeneration of list strata over time.

1. INTRODUCTION

The Agriculture Division of Statistics Canada conducts the Agriculture Enumerative Survey (AES) every July. The AES is a multi-purpose survey based on an area sample producing estimates for crops, livestock and expense items for all provinces but the Prairies. A recurring problem with the survey has been that the sample size allocated to smaller provinces has been insufficient to produce good provincial estimates. It was decided, therefore, to test multiple frame sampling in one of these provinces to determine whether this technique could increase the efficiency of estimates and to study the operational problems associated with the technique. The province chosen for testing was New Brunswick.

This was our first experience with multiple frame sampling and as such it was a learning experience. This paper will present the results of the New Brunswick test which, although successful in reducing sampling errors, pointed out problem areas. These problems as well as proposed solutions will be discussed.

---

[1] B. Armstrong, Institutional and Agriculture Survey Methods Division, Statistics Canada.

2. SAMPLE DESIGN FOR THE 1978 NEW BRUNSWICK TEST

A multiple frame survey is one which employs two or more sampling frames to produce estimates for a specific survey population. In the 1978 New Brunswick test an area and a list frame were combined to produce estimates for agricultural items. Use of the area frame was essential to ensure complete coverage of the population. The list frame was introduced to improve the efficiency of the sample design.

The AES, in its present form employs an extreme type of multiple frame sampling. The AES area sample is supplemented by a group of very large farms (large with respect to some key items) taken from the updated 1976 Census of Agriculture list and included in the sample with probability 1. This group of farms, referred to as specified farms, is included in the sample as a separate, complete enumeration stratum for two reasons. Since the AES is a probability sample survey, the values of items for farms in the sample are blown up to represent a larger group of farms. If these specified farms were not identified prior to the survey and were by chance picked up in the area sample the resulting estimates would be blown up out of proportion since they are not typical farms. On the other hand, since these farms contribute a significant amount to the provincial totals, their chance exclusion from the sample would result in lower estimates. In either case--chance inclusion or exclusion--the estimates would vary considerably and so specified farms are included with probability 1. This feature of the AES sample design was retained in the multiple frame test also, as described below.

a) Stratification of the List Frame

The list frame for the New Brunswick multiple frame test, was the updated 1976 Census of Agriculture list. Very small farms were excluded from the Census list and the remaining farms formed the list sampling frame. This frame was stratified as shown in Table 1 and a simple random sample selected from each stratum. Starting with

stratum 1 and continuing sequentially, a farm was assigned to the first stratum for which it met the stratum criterion. As we shall see later, this method of stratifying the list proved to be inefficient for the estimation of one of the stratifying items.

TABLE 1: Stratification of the List Frame

Stratum	Stratum Definition	Population Size	Sample Size
1	Specified farms as defined in AES	69	69
2	Total sales > $75,000	200	88
3	Total potatoes > 50 acres	261	60
4	Total cattle > 40	448	60
5	Total pigs > 30	83	30
6	Total potatoes > 10 acres	204	15
7	Total cattle > 25	858	30
8	Total pigs > 8	84	15
		2,207	367

Sample allocation to strata was based on a trial and error method. The allocation ultimately chosen was the one which gave the best combination of coefficients of variation for the three key items (i.e. potatoes, cattle, pigs) in New Brunswick.

b) Stratification of the Area Frame

The design of the area sample was the same as it had been since the last redesign in 1974. Enumeration areas (EA's) as defined in the 1971 Census of Agriculture were the first stage sampling units. EA's were stratified in a manner similar to the stratification of the list frame using Census data (summarized at the EA level) as shown in Table 2. A stratified replicated random sample of EA's was then selected. Again, sample allocation was based on trial and error with the chosen allocation giving the best combination of coefficients of variation for key items in the province.

Selected EA's were divided into roughly equi-sized pieces of land or segments. These segments became the second stage sampling units. A sample of one or more segments was selected from each selected EA

depending on the size of the EA. All operators with land within selected segments were enumerated.

TABLE 2: Stratification of Area Frame

Stratum	Stratum Definition	Population Size	Sample Size	No. of Replicates	No. of EA's per Replicate	Sample Size
	(EA Level)	(# EAs)	(# EAs)			(# Farms)
1	Total chickens ≥ 25,000	25	12	6	2	29
2	Total potato acs. ≥ 6,000	30	30	15	2	212
3	Total pigs ≥ 400	21	14	7	2	32
4	Total cattle ≥ 500	62	20	10	2	92
5	X ≥ 47	56	14	7	2	42
6	X ≥ 14	136	18	9	2	30
7	Remaining EA's	219	16	16	1	29
8	OLD non-agricultural EA's	221	6	6	1	0
9	NEW non-agricultural EA's	57	2	2	1	0
						466

NOTE: X was a conglomerate variable constructed for the province using a combination of livestock and crop items. For those EA's which displayed no dominant agricultural characteristic the variable X was used to form strata containing EA's with similar levels of agricultural activity.

X = 20,000* [(ratio of the number of livestock in the EA to the total number of livestock in all agricultural EA's in the province) + (ratio of cropland area in the EA to total cropland area of all agricultural EA's in the province)].

As 1978 was our first test of multiple frame, an adequate sample had to be allocated to the area frame in order to secure the regular AES area sample estimates should there be a problem with the multiple frame estimates. Fortunately, due to a F.L.I.P. grant (Federal Labour Intensive Program) from the federal government, the sample allocation to New Brunswick was increased to such an extent that we were able not only to add a list sample to the existing area sample but to actually increase the area sample above its 1977 level.

3. ESTIMATION PROCEDURES

From a paper by Hartley (see reference [1]), a single multiple frame estimator may be obtained from two survey frames (in this case, area and list) by adding a combined area and list sample estimate for the overlap domain (i.e. the portion of the population covered by the list frame) to the area sample estimate for the non-overlap domain (i.e. the portion of the population not covered by the list frame). This Hartley estimator is

$$\hat{Y}_H = \hat{Y}_{NOL} + q\,\hat{Y}_{OL} + p\,\hat{Y}_L$$

where $\hat{Y}_{NOL}$ = area frame estimate for the non-overlap domain,

$\hat{Y}_{OL}$ = area frame estimate for the overlap domain,

$\hat{Y}_L$ = list frame estimate,

p = weight given to list frame estimate,

q = weight given to area frame estimate

and $p+q = 1$.

An optimum value of p based solely on variance minimization can be determined (see Appendix 1 for derivation of p_{opt}). Hartley derives an optimum value for p which minimizes the multiple frame variance estimator with respect to a cost function which depends on the unknown sample sizes for both the list and the area frames. These sample sizes are then optimized along with p. In the New Brunswick test, however, the sample allocations to the list and area frames were determined based on other considerations. The area sample had to be sufficiently large to produce adequate area sample estimates. Thus, the size of the area sample was fixed before the list frame was introduced. The sample size allocated to the list frame was then set arbitrarily. As well, in the AES there is virtually no difference in the mapping and enumeration costs of list and area sample farms since data collection for both types of farms is done by interview. For these reasons, neither the cost function nor the optimum allocation of the sample between the two frames **was** used to determine the optimum values of p.

The formula for the variance estimate of the Hartley estimator is given in Appendix 1.

The multiple frame screening estimator is a special case of the Hartley estimator where $p=1$ and $q=0$. In this case only the list sample is used to estimate for the overlap domain. The area sample estimates only for the non-overlap domain. Using the notation defined earlier, the screening estimator takes the form

$$\hat{Y}_S = \hat{Y}_L + \hat{Y}_{NOL}.$$

The variance of the screening estimate is simply

$$\text{Var}\ (\hat{Y}_S) = \text{Var}\ (\hat{Y}_L) + \text{Var}\ (\hat{Y}_{NOL}).$$

The regular area sample estimator is, again, a special case of the Hartley estimator where $p=0$ and $q=1$. Thus, the area sample estimator is

$$\hat{Y}_A = \hat{Y}_{OL} + \hat{Y}_{NOL}$$

with variance

$$\text{Var}\ (\hat{Y}_A) = \text{Var}\ (\hat{Y}_{OL}) + \text{Var}\ (\hat{Y}_{NOL}) + 2\ \text{Cov}\ (\hat{Y}_{OL}, \hat{Y}_{NOL}).$$

For details of the formulae for area sample estimators of totals and variances see Appendix 2.

4. 1978 NEW BRUNSWICK ESTIMATES

The multiple frame screening estimates for the 1978 New Brunswick test are presented in Table 3. Comparing regular AES area frame estimates with multiple frame estimates, there were two important observations to note. First of all, coefficients of variation for multiple frame estimates were significantly lower than for area frame estimates--often by as much as 50%. The second observation was that the level of multiple frame estimates appeared to be generally higher than for

area frame estimates. In only 4 out of the 21 estimates displayed in Table 3 were multiple frame estimates lower than the area frame estimates.

The first observation needs little comment. The list frame is a more efficient sampling frame and we therefore expected the coefficients of variation to decrease sharply with the introduction of a list sample.

It was with the second observation with which there was the most concern. However, although there did appear to be a tendency for multiple frame estimates to be higher than the area frame estimates, it is interesting to note that the same tendency could be seen with the published estimates. Published estimates are compiled by subject matter experts taking into account estimates from all their sources (of which the AES is one). Of the 17 cases where the multiple frame estimate was higher than the area frame estimate, 11 of the published estimates were also higher than the area frame estimate. In all the 4 cases where multiple frame estimates were lower than the area frame estimates, the published estimates were lower than the area frame estimates as well. Thus the "level" problem of multiple frame estimates does not appear as extreme taking this into account although we shall be investigating it further in Section 10.

TABLE 3: New Brunswick Estimates - 1978

Item	Area Frame Estimate	C.V.	Multiple Frame Screening Estimate	C.V.	Published Figure
Total area (acres).....	1,090,235	8.5	1,271,419	6.0	1,090,200
Potatoes (acres)	63,355	15.4	70,318	7.2	58,000
Mixed grains (acres) ..	3,945	46.5	5,077	37.8	6,800
Oats (acres)	37,357	17.2	45,771	9.4	42,000
Barley (acres)	5,210	26.6	5,946	19.5	7,200
Tame hay (acres)	158,628	12.0	197,167	7.5	180,000
Spring wheat + Winter wheat (acres) ..	7,903	-	8,405	-	9,600
Corn for grain (acres).	380	31.9	510	33.8	500
Total crops (acres) ...	298,238	9.8	356,746	5.4	298,200
Improved land for pasture (acres) ...	97,987	13.3	119,155	10.0	-
Total cattle	109,350	12.0	118,844	5.4	113,000
Milk cows > 2 yrs	35,277	20.2	34,801	9.6	31,000
Beef cows > 2 yrs	19,300	13.8	23,815	11.8	23,000
Dairy heifers (1-2 yrs)	11,071	23.4	9,804	11.0	7,600
Bulls	2,600	16.6	2,743	16.1	2,700
Steers	7,200	12.1	8,821	11.1	8,200
Calves	25,500	11.1	29,365	11.1	29,700
Total pigs	47,610	18.8	60,925	12.1	43,000
Sows & gilts	10,205	44.1	7,741	14.1	5,000
Boars	700	38.5	587	17.8	400
Market pigs (<3 mon) ..	21,200	16.9	29,148	14.1	18,000

5. COMPARISON OF SCREENING AND HARTLEY MULTIPLE FRAME ESTIMATES

Table 4 shows the area frame, screening and Hartley estimates for the four key items in New Brunswick. As well, the p and q values of the Hartley estimate are given.

TABLE 4: Comparison of Different Estimates for N.B., 1978 (C.V.'s in brackets)

	Area Frame Estimate	Standard Error	Screening Estimate	Standard Error	Hartley Estimate	Standard Error	p	q
Total Area (acs)	1,090,235 (8.5%)	92,730	1,271,419 (6.0%)	75,726	1,217,572 (5.9%)	71,487	0.70	0.30
Potato acres .	63,355 (15.4%)	9,788	70,318 (7.2%)	5,050	69,900 (7.2%)	5,023	0.94	0.06
Total Cattle .	109,351 (12.0%)	13,124	118,844 (5.4%)	6,410	117,577 (5.2%)	6,159	0.75	0.25
Total pigs ...	47,610 (18.8%)	8,960	60,925 (12.1%)	7,365	55,581 (10.4%)	5,751	0.59	0.41

Results for total area, potato area, and total cattle are as expected. Standard errors have been reduced slightly by using the Hartley estimate and levels of the Hartley and screening estimates are comparable. It is the total pig estimates which are interesting. The Hartley estimate is lower than the screening by 5,000. The standard error for the Hartley estimate is also substantially lower, with a coefficient of variation of 10.5% (as opposed to 12.1% for the screening estimate). The reason for this lies with the list estimate for total pigs. Since the variance of the list estimate was relatively high (although still about half that of the area estimate for the overlap domain), the result was that a lower weight of p was assigned to the list estimate of the overlap domain and consequently gains in efficiency were realized using the combination of estimates for the overlap domain in the Hartley estimate. The value of p for total pigs was 0.59.

6. LIST ESTIMATES - COMPARISON WITH 1976 CENSUS

For multiple frame sampling to be effective in producing good estimates, it is essential that the list sample provide a good estimate for that portion of the population covered by the list frame. As a first step in evaluating list estimates, 1978 list estimates were compared by stratum with the corresponding totals from the updated 1976 Census.

TABLE 5: Comparison of 1976 Updated Census Totals and 1978 List Estimates, N.B.

Stratum	Total Area (acs)		Potato Acres		Total Cattle		Total Pigs	
	1976	1978	1976	1978	1976	1978	1976	1978
1	50,954	68,888	2,123	3,967	13,416	13,359	18,443	19,427
2	99,644	93,498	17,721	19,148	7,145	5,082	4,191	1,809
3	87,284	102,155	26,243	25,461	4,005	2,462	561	3,306
4	203,822	245,601	678	866	39,186	37,438	3,293	12,484
5	19,518	19,973	168	357	1,895	1,494	9,278	9,907
6	43,453	58,303	6,193	8,418	2,790	3,006	268	1,210
7	242,836	229,229	237	86	34,303	26,255	1,745	4,976
8	13,964	12,634	39	34	912	1,058	1,228	885
Total	761,475	830,281	53,402	58,336	103,652	90,154	39,007	54,005
Coefficient of Variation	-	5.5%	-	6.3%	-	4.6%	-	12.9%

Table 5 shows corresponding figures from 1976 and 1978 for four key items in New Brunswick. As was mentioned in the last section, the list estimate for total pigs had a high coefficient of variation. This is not surprising since at the design stage, pigs were given the lowest priority of all stratification variables. This low priority may have been responsible for the large difference between the census total and the list estimate. Notice that the cattle strata (strata 4 and 7) contributed 17,460 to the total pig estimate while the census total for these two strata was only 5,038. Since cattle had a higher priority than pigs at stratification, if a farm met the criteria for both the large cattle and the large pig strata, it was put in the large cattle stratum. Thus, the resulting stratum was homogeneous for total cattle but not for total pigs. The result was that, because

some large pig farms were picked up in the sample, the stratum produced an inflated estimate for total pigs. For future surveys it would be wise to consider giving pigs an equal priority with cattle.

The list estimate for potato acreage was 58,336. The portion of the census list used as a list frame in New Brunswick accounted for 96.18% of all potato area in the province in 1976. Thus, if we had had only the list sample from which to estimate, results would have been similar to the 1978 New Brunswick potato objective yield survey (a survey based on a list sample only). The fact is, however, that we had an area sample as well which showed that area sample farms not found on the list frame contributed 11,982 acres of potatoes to the provincial estimate. Farms which were on the census list but not on the reduced list frame used in New Brunswick accounted for 4,680 acres. This means that the estimate coming from farms which were not found on the census list was 7,302 acres. It is not surprising, therefore, that multiple frame estimates were higher than the New Brunswick potato objective yield survey estimates, the latter being based solely on a list frame.

7. CONDITION OF LIST FRAME

The main problem in using any list frame to select a sample is that such frames become out of date very quickly. The census data upon which stratification in New Brunswick was based was two years old when the survey was run. Table 6 shows for list sample farms to what stratum each farm was assigned before the survey and to what stratum it should have been assigned based on survey data.

TABLE 6: List Sample Stratum Changes 1976-1978

1976 Stratum	1978 Stratum										Total
	1	2	3	4	5	6	7	8	9*	Out of Business	
1	51	8	1	9	-	-	-	-	-	-	69
2	2	66	9	2	1	-	-	-	6	2	88
3	1	11	35	-	-	4	-	-	-	9	60
4	2	3	1	52	-	-	-	-	-	2	60
5	1	-	1	4	14	-	5	1	2	2	30
6	-	-	4	-	-	6	-	1	2	2	15
7	-	-	-	8	1	-	11	-	5	5	30
8	-	-	-	-	2	-	-	4	8	1	15
Total .	57	88	51	75	18	10	16	6	23	23	367

*Stratum 9 refers to farms which no longer meet the criteria of any list stratum.

As is shown in Table 6, the strata for large farms (strata 1-5) tended to be more stable than the lower strata thereby reducing or eliminating the benefits of stratification for these lower strata. Thus, it may be advisable to put stricter limits on strata to be included in the list frame. The smaller sized farms are the ones which can more readily "jump" strata so that the area frame would estimate almost as well for these farms. This would also leave the entire list sample to estimate more efficiently for the larger sized farms.

8. REMOVING MEDIUM STRATA FROM THE LIST

Eliminating a stratum from list frame coverage changes the multiple frame sample in two ways. First, list sample farms in that stratum do not contribute to the list estimate. Secondly, area sample farms which overlap with that stratum in the population frame (and were, therefore, part of the area overlap domain) are not part of the non-overlap domain. For the screening estimate described in Section 3, then, the number of farms contributing to the list portion of the estimate decreases while the number of farms contributing to the area portion of the estimate increases.

Table 7 shows the effect on multiple frame (screening) estimates of removing medium-sized strata (strata 6,7,8) from list frame coverage. Coefficients of variation for all items increased only slightly (over the full multiple frame estimates) but this increase was significant considering that the sample size increased as well. Coefficients were still lower than for area frame estimates. Also worth noting is the fact that allowing the area frame to estimate for these strata reduced the level of resulting estimates for all items. Therefore the area frame estimates were lower than the list frame for these list strata. This may have been due to the problem of out-of-date stratification in the lower list strata as mentioned in the previous section. Since the area frame appeared to estimate better than the list for lower strata the conclusion reached in Section 7 is reinforced here, i.e. the medium-sized strata should be removed from list frame coverage for the 1979 survey.

TABLE 7: Estimates for Reduced List Frame Coverage (C.V.'s in brackets)

Item	Area Frame Estimate	Full Multiple Frame Estimate	Removing Strata 6, 7, 8 from List Frame
Total area (acres)	1,090,235 (8.5%)	1,271,419 (6.0%)	1,220,732 (6.4%)
Potato acres	63,355 (15.4%)	70,318 (7.2%)	66,600 (7.2%)
Total Cattle	109,350 (12.0%)	118,844 (5.4%)	114,483 (6.5%)
Total Pigs .	47,610 (18.8%)	60,925 (12.1%)	57,082 (12.1%)
Sample Size.	535	569	591

9. ELIMINATING STRATA BY COMMODITIES

Table 8 shows what happened as potato strata (3 and 6), cattle strata (4 and 7), and pig strata (5 and 8) were, in turn, dropped from list frame coverage. As strata based on a certain item are dropped from the list frame one would expect the efficiency of the estimate for that item to drop, since the area sample is now estimating almost entirely for it. As shown in Table 8 this was, in fact, what happened. As potato strata were removed the coefficient of variation for potatoes increased from 7.2 to 11.1; as cattle strata were removed the coefficient for cattle increased from 5.4 to 11.7; and as pig strata were removed the coefficient for total pigs increased from 12.1 to 16.6.

What is interesting to note in Table 8 is what happened to the estimate for total pigs as the cattle strata (4 and 7) were removed from list coverage. By allowing the area sample to estimate for the cattle strata, the estimate for total pigs was reduced to the level of the 1978 area frame estimate. As well, the coefficient of variation for this estimate was less than for both the area frame and the full multiple frame estimate. This tended to support the suspicion raised in Section 6 that the list sample for the cattle strata over-estimated for total pigs. It also reinforced the recommendations that, for stratification purposes, total pigs should be given equal or higher priority with total cattle for future design of list frames.

TABLE 8: Eliminating List Strata for Certain Commodities (C.V.'s in brackets)

Item	Area Frame Estimate	Full Multiple Frame Estimate	Eliminate Strata 3, 6 (Potatoes)	Eliminate Strata 4, 7 (Cattle)	Eliminate Strata 5, 8 (Pigs)
Total area .	1,090,235	1,271,419	1,216,582	1,145,174	1,263,446
(acres)	(8.5%)	(6.0%)	(6.0%)	(8.1%)	(6.0%)
Potato acres	63,355	70,318	62,191	70,151	69,944
	(15.4%)	(7.2%)	(11.1%)	(7.1%)	(7.2%)
Total Cattle	109,350	118,844	118,585	106,582	119,271
	(12.0%)	(5.4%)	(5.2%)	(11.7%)	(5.6%)
Total Pigs .	47,610	60,925	56,868	48,100	60,945
	(18.8%)	(12.1%)	(12.0%)	(9.7%)	(16.6%)
Sample size.	535	569	591	585	537

10. OVERLAP DETERMINATION

The high level of multiple frame estimates indicated that there could be a problem with the determination of overlap between the area sample and list frame. Recall that only area sample farms which are not found on the list frame contribute to the area portion of the multiple frame estimate. If area sample farms which appear on the list frame are not identified as such, then resulting multiple frame estimates will be inflated.

To check on the overlap determination, a list of area sample farms not matched to the list frame at head office was sent to the regional office to verify that they were true "non-matches" to the list frame. The list was returned with comments indicating that several farms were on the list frame but were now operating under different names. While the regional office assumed that we were matching farms (i.e. pieces of land) we were in fact matching farm operators. If an area sample farm appeared on the list frame under a different operator's name, it was next to impossible for head office to identify these as being the same farm. Therefore we had to be content with matching farm operators rather than pieces of land.

However, after this confusion with regional office over "farms" versus "farm operators", it occurred to us that the same confusion had probably arisen with enumerators of list sample farms. Again, since it is a farm operator that is selected from the list frame, the list sample farm should be classified as being out of business if the operator is no longer the same. As this had never been explicitly told to enumerators they tended not to make this distinction. Farms were enumerated as usual even if the farm operator had changed. Thus, it was necessary to go through the list questionnaires to pick out all such farms (of which there were 12). Table 9 gives the revised estimates produced by removing the contribution of these farms from multiple frame estimates.

TABLE 9: Revised Multiple Frame Estimates

Item	Original Multi-frame Estimate	Revised Multi-frame Estimate	Weighted Estimate	Published Estimate
Total area .	1,271,419 (6.0)	1,241,888 (6.2)	1,090,235 (8.5)	1,090,200
Potatoes ...	70,318 (7.2)	66,479 (7.5)	63,335 (15.4)	58,000
Total Cattle	118,844 (5.4)	115,334 (5.6)	109,350 (12.0)	113,000
Total Pigs .	60,925 (12.1)	58,388 (12.3)	47,610 (18.8)	43,000

The revised multiple frame estimates do show improvement over the original estimates in terms of the level of estimates. Thus much of our "overlap" problem was largely definitional. More explicit instructions and clearer definitions of out of business list farms will be given to enumerators next year. Further work, however, will have to be done to verify the quality of overlap determination for the 1979 survey.

11. CONCLUSIONS AND RECOMMENDATIONS

A great deal of information was gained from the 1978 New Brunswick test. Generally speaking, the test ran smoothly and results are encouraging. It appears that multiple frame sampling is a viable technique for collecting data and producing estimates in a province such as New Brunswick. It has already been decided that use of multiple frame will expand to all three Maritime provinces for the 1979 survey. It will be a year for further testing of multiple frame since once again area frame estimates will be produced in addition to multiple frame estimates.

In light of the analysis outlined in this paper, the following recommendations were made and adopted for the 1979 survey.

1. Since data are to be collected for overlap area farms in 1979. Hartley estimates as well as screening estimates will be produced. Although for most items the gain in efficiency when using the Hartley estimates will be slight, if the data are available, the additional computations necessary should be made. This will prove especially worthwhile for items for which the list estimate is poor for it is with these items that the Hartley estimate shows the greatest improvement.

2. The medium sized strata will be dropped from the list frame. List estimates for these strata are not as good as for the larger strata. It has been shown that a large number of sample farms selected in these strata have changed strata by survey time, thus reducing the effectiveness of stratification. As well, small to medium sized list frame farms with limited data available may be harder to match with area sample farms. Thus, the crucial step of overlap determination could be endangered by including these farms in the list frame.

3. At the stratification stage, pigs are to be given equal priority with cattle when defining list strata. There is evidence from 1978 data that giving cattle priority has caused problems with list estimates for pigs. Equal priority will be given to cattle and pigs through use of multiple (or deep) stratification techniques.

4. Interviewers will attempt to determine overlap between the area sample and the list frame in the field in 1979. They will carry with them a copy of the list frame and, after every interview with an area frame sample farm, will record whether or not this farm is on the list frame. It is hoped that interviewers will be able to make this determination more accurately than head office personnel were able to do in 1978. For 1979, however, the entire area sample/list frame match will be checked at head office to evaluate just how well interviewers were able to perform this step.

In addition, a question has been added to the 1979 questionnaire which will help the interviewer determine whether the area sample farm is likely to be found on the list frame. The question is

"Was this farm operating under the present name at June 1, 1976?

If the answer to this is no, the farm is automatically a non-overlap farm; if yes, the farm should have been included on the census list (but not necessarily the list frame) and the list frame would therefore be carefully checked for a match.

In conclusion, it is felt that multiple frame testing has certainly been worthwhile. We are continuing the testing in 1979 and if the changes to the 1979 survey (as listed above) are successful, we will be able to significantly improve our estimates in future surveys.

12. ACKNOWLEDGEMENT

The author would like to thank the referee for helpful comments and suggestions.

RESUME

La question étudiée dans le présent document est l'estimation des diverses variables agricoles selon la méthode des listes multiples. La liste des secteurs de dénombrement comprend intégralement la liste complémentaire. La stratification de la liste complémentaire et de la liste des secteurs de dénombrement se fondent sur des critères différents. De façon générale, la méthode des listes multiples présente certains avantages sur le plan de la variance par rapport à la liste des secteurs de dénombrement. Cependant, une analyse plus poussée révèle que la liste complémentaire comporte des lacunes, telles que la méthode de stratification et la dégénération des strates dans le temps.

REFERENCE

[1] Hartley, H.O., "Multiple Frame Surveys", Proceedings of the Social Statistics Section of the American Statistical Association meeting, Minneapolis, Minnesota, (1962).

APPENDIX 1

Multiple Frame Formulae

a) Notation

$\hat{Y}_L$ = estimate for list frame population from list sample.

$\hat{Y}_A$ = area sample estimate for entire population.

$\hat{Y}_{OL}$ = area sample estimate of list frame population (overlap domain).

$\hat{Y}_{NOL}$ = area sample estimate of population not covered by list frame (non-overlap domain).

p = weight given to list frame estimate (for Hartley estimate).

q = weight given to area frame estimate of list frame population.

$p + q = 1$.

b) Area Frame Estimate

$$\hat{Y}_A = \hat{Y}_{OL} + \hat{Y}_{NOL} .$$

$$\text{Var}(\hat{Y}_A) = \text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_{NOL}) + 2\,\text{Cov}(\hat{Y}_{OL}, \hat{Y}_{NOL}) .$$

c) Multiple Frame Screening Estimate

$$\hat{Y}_S = \hat{Y}_L + \hat{Y}_{NOL} .$$

$$\text{Var}(\hat{Y}_S) = \text{Var}(\hat{Y}_L) + \text{Var}(\hat{Y}_{NOL}) .$$

d) Hartley Multiple Frame Estimate

$$\hat{Y}_H = \hat{Y}_{NOL} + q\,\hat{Y}_{OL} + p\,\hat{Y}_L .$$

The area frame estimate is a special case of the Hartley estimate when $p = 0$ and $q = 1$.

The screening estimate is a special case of the Hartley estimate when $p = 1$ and $q = 0$.

The Hartley estimate may be written

$$\begin{aligned}\hat{Y}_H &= \hat{Y}_{NOL} + (1-p)\ \hat{Y}_{OL} + p\ \hat{Y}_L \\ &= \hat{Y}_{NOL} + \hat{Y}_{OL} - p\ \hat{Y}_{OL} + p\ \hat{Y}_L \\ &= \hat{Y}_A + p(\hat{Y}_L - \hat{Y}_{OL}).\end{aligned}$$

The variance of this estimate is

$$\begin{aligned}\text{Var}\ (\hat{Y}_H) &= \text{Var}(\hat{Y}_{NOL}) + q^2\text{Var}(\hat{Y}_{OL}) + p^2\text{Var}(\hat{Y}_L) + 2\ q\ \text{Cov}(\hat{Y}_{NOL},\hat{Y}_{OL}) \\ &= \text{Var}(\hat{Y}_{NOL}) + \text{Var}(\hat{Y}_{OL}) - 2\ p\ \text{Var}(\hat{Y}_{OL}) + p^2\text{Var}(\hat{Y}_{OL}) + p^2\text{Var}(\hat{Y}_L) \\ &\qquad + 2\ \ \text{Cov}(\hat{Y}_{NOL},\hat{Y}_{OL}) - 2\ p\ \text{Cov}(\hat{Y}_{NOL},\hat{Y}_{OL}) \\ &= [\text{Var}(\hat{Y}_{NOL}) + \text{Var}(\hat{Y}_{OL}) + 2\ \text{Cov}(\hat{Y}_{NOL},\hat{Y}_{OL})] \\ &\quad - 2\ p[\text{Var}(\hat{Y}_{OL}) + \text{Cov}(\hat{Y}_{OL},\hat{Y}_{NOL})] + p^2[\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)] \\ &= \text{Var}(\hat{Y}_A) - 2\ p\ [\text{Cov}(\hat{Y}_A,\hat{Y}_{OL})] + p^2[\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)].\end{aligned}$$

The optimum value of p is now determined.

$$\frac{\delta\ \text{Var}(\hat{Y}_H)}{\delta\ p} = -\ 2\ \text{Cov}(\hat{Y}_A,\hat{Y}_{OL}) = 2\ p\ [\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)] = 0\ .$$

$$\therefore\ p_{opt} = \frac{\text{Cov}(\hat{Y}_A,\hat{Y}_{OL})}{[\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)]}\ .$$

Using this value of p the variance of the Hartley estimate becomes:

$$\begin{aligned}\text{Var}\ (\hat{Y}_H) &= \text{Var}(\hat{Y}_A) - 2\ p_{opt}\ [\text{Cov}(\hat{Y}_A,\hat{Y}_{OL})] + p^2_{opt}\ [\text{Var}(\hat{Y}_L) + \text{Var}(\hat{Y}_{OL})] \\ &= \text{Var}(\hat{Y}_A) - 2\ p^2_{opt}\ [\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)] + p^2_{opt}\ [\text{Var}(\hat{Y}_L + \text{Var}(\hat{Y}_{OL})] \\ &= \text{Var}(\hat{Y}_A) - p^2_{opt}\ [\text{Var}(\hat{Y}_{OL}) + \text{Var}(\hat{Y}_L)].\end{aligned}$$

APPENDIX 2

Area Sample Estimator

As shown in Table 2 (Section 2), selection of first stage sampling units in the AES is replicated. Data from all EA's within a replicate are blown up to the stratum level. The estimator for stratum h from replicate k is

$$\hat{y}_{hk} = \frac{M_h}{m_h} \sum_{i=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{j=1}^{n_{hi}} y_{hij},$$

where M_h = the number of EA's in stratum h,

m_h = the number of EA's selected per replicate for stratum h,

N_{hi} = the number of segments in the ith selected EA of stratum h,

n_{hi} = the number of selected segments in the ith selected EA of stratum h,

y_{hij} = data value for the jth segment of the ith EA in stratum h.

Usually a farm enumerated within a segment has part of its land lying inside the segment and part outside. For such farms, data values are reduced to the segment level by applying a weight equal to the ratio of the farm's land inside the segment (excluding woodland) to total land operated on the farm (excluding woodland). It is this 'weighted" data value for each farm within a segment which is summed to give the segment total y_{hij}.

Having obtained estimates for all replicates, the stratum estimate can be calculated and is

$$\hat{y}_h = \frac{\hat{y}_{h1} + \hat{y}_{h2} + \dots \hat{y}_{hr_h}}{r_h}$$

where r_h = number of replicates in stratum h.

To obtain provincial estimates, stratum estimates are summed

$$\hat{Y}_A = \sum_{h=1}^{H} \hat{Y}_h$$

where H = number of strata in the province.

Estimates of variance take the form

$$\hat{V}(\hat{Y}_A) = \sum_{h=1}^{H} \frac{1}{r_h(r_h-1)} \{(\hat{Y}_{h1} - \hat{Y}_h)^2 + \ldots \quad (\hat{Y}_{hr} - \hat{Y}_h)^2\}.$$

CANADIAN VICTIMIZATION SURVEYS: A REPORT ON PRETESTS IN EDMONTON AND HAMILTON

Gary Catlin and Susan Murray[1]

This article presents the methodology and analysis of two major pretests undertaken in order to compare the effectiveness of different interviewing methods and to assess the feasibility of collecting information which would meet Victimization Survey information requirements.

1. INTRODUCTION

The Department of the Solicitor General of Canada contracted with Statistics Canada to develop a methodology for conducting Victimization Surveys in Canada. The Research Division of the Solicitor General and the Special Surveys staff of Statistics Canada jointly participated in this research program. A Victimization Survey methodology is designed to produce data related to four main objectives:

1. the extent and distribution of selected crimes;
2. the impact of selected crimes;
3. the risk of criminal victimization;
4. indicators of criminal justice system functioning.

These data will be useful to those directly involved in efforts to contain or reduce criminal activity. More generally, victimization data will also be useful to policy makers and evaluators who are concerned with crime and its efforts upon society, and to social scientists interested in advancing the state of knowledge about crime and the criminal justice system.

[1] Gary Catlin, Special Surveys Group and Susan Murray, Census and Household Survey Methods Division, Statistics Canada.

The crimes for which data are collected are break and enter, theft, robbery, assault, rape, motor vehicle theft and vandalism. Much of the data available concerning crime and the criminal justice system has been based on official police statistics (Uniform Crime Reporting or U.C.R. statistics). However, victimization surveys are capable of providing data on crimes not reported to the police as well as collecting more information than is presently available on those crimes reported to the police.

As a consequence of the statistical rarity of criminal victimization, it is necessary for victimization surveys to question a large number of persons in order to obtain sufficient information to allow for meaningful analyses. Given the necessity for large sample sizes, cost effectiveness, while maintaining data quality, was the primary concern of the methodological studies reported here.

These studies addressed cost effectiveness in two ways:

1. they assessed the feasibility of collecting information related to the four major objectives; and
2. they compared the effectiveness of personal and telephone interviewing methods.

The initial pretest was conducted in Edmonton in May 1977 and had, as its primary purpose, a comparison of personal and telephone interviewing techniques. The second pretest was conducted in Hamilton during February 1978 to test the ability of the revised questionnaire to generate reports of crime and to refine the interviewing procedures. This report presents the findings of these studies. The Edmonton pretest is presented in detail and a summary of the Hamilton pretest is included.

2. EDMONTON REVERSE RECORD CHECK: OBJECTIVES

2.1 Data Collection Methodologies

The primary objective of the Edmonton Reverse Record Check was to assess the feasibility of telephone interviewing. The following criteria were used to compare telephone and personal interviewing:

(a) Response Rate

The rate and type of non-interview is essential in evaluating the telephone and personal interview methods. It may also be important to discover what specific items of information are refused or considered sensitive by the respondent.

(b) Hit Rate

The hit rate refers to the proportion of incidents selected from the police files which were subsequently reported by the respondent during the survey interview. The accuracy of the information reported will also affect the evaluation of the two interviewing methods.

(c) Cost

The impetus for investigating the feasibility of telephone interviewing is an expected savings of approximately 70%. Therefore, if the telephone and personal interviewing methods are comparable, based on all other criteria, the more cost efficient method would be employed.

(d) Telescoping

It is also important to measure the extent of telescoping, that is, the respondents' misplacing of events in time.

(e) Unfounded Crimes

An unfounded crime is an incident which has been reported to the police but which the police have determined, due to subsequent events, does not constitute a crime. A sample of such incidents was selected to discover whether these events would be reported and, if reported, how the survey document would classify them.

2.2 Evaluation of Questionnaire Design

The second broad objective of the Edmonton Reverse Record Check was to evaluate the ability of the questionnaire to supply the necessary information required in a victimization survey. Identical questionnaires were employed for both the personal and telephone methodologies. Two separate forms were utilized. The first, the screen questionnaire, was administered to all respondents and consists of items on selected attitudes, behaviour and demographic characteristics. As well, it includes a series of questions designed to elicit incidents of crime which had been committed against the respondent or, for certain crimes, his/her household. These questions serve to facilitate the recall and reporting by the respondent of crimes that occurred within the reference period. The second, the crime incident report, was then filled out for each incident reported in the screen question section of the questionnaire after the entire screen questionnaire had been completed. The crime incident report includes questions to determine the social and physical context of the incident and its impact on the respondent as well as the information required to classify the incident into crime categories.

Primarily, it was necessary for the questionnaire to elicit reports of incidents of crime. Specifically, it was important to know which probes were most effective. The interviewing also quickly illuminated problems or difficulties which were caused by the design of the questionnaire.

The objectives of a victimization survey require that some interviews of long duration be conducted. The minimum length of an interview was approximately ten minutes. Given that conventional wisdom stipulates that it is inadvisable to conduct long interviews by telephone, it was necessary to determine if the length of interview caused some respondents to terminate prior to completing the interview.

An eventual survey, designed to measure among other things the incidence of crime, must elicit sufficient information to categorize reported victimizations according to approximate U.C.R. and Criminal Code definitions. Of course, precise conformity is not possible. Further, classifying by U.C.R. definitions is not the primary objective with regard to coding the victimizations, but merely the minimal information that should be available. Two victimizations that U.C.R. may code as 'robbery' could be classified more meaningfully as 'a personal attack with theft' or 'personal attack, involving a weapon, with theft'. The additional information collected will allow the analyst to define victimizations in ways that may be much more meaningful for planning and evaluating purposes.

The objective of the reverse record check, in terms of classification, was to determine the comparability between the classification of the crime by the survey instrument and the classification of the incident according to the police.

3. STUDY DESIGN

3.1 Police File Sample

The reverse record check technique involves selecting samples of known victims from police files. These people are then interviewed to determine the degree to which they report the criterion event (i.e., the incident on the basis of which the respondent was selected) and its associated characteristics to the interviewer. Thus by assigning a sample of cases selected from police files to both interviewing methods, it is possible to assess the relative success of each data collection method.

With the cooperation of the Edmonton City Police Department, a sample of victims of crimes was selected from the offence reports in the police files. Each individual chosen was listed as a complainant or victim of one of the crimes included in the victimization survey, even though in some cases the incident was not reported to the police by that person. There were three distinct subsamples chosen from the files - the reference sample, the telescope sample and the unfounded sample.

Police files do not provide a perfect criterion measure since, of course, they include only crimes reported to the police. Such crimes may be more salient to respondents when interviewed later and may be more easily remembered than the types of incidents not reported. In addition, the errors and omissions which occur in police recording practices make the use of police files less than perfect. On the other hand, police occurrence reports contain a number of items which can be checked against the information provided by survey respondents. These include the nature of the loss suffered, the nature of any injury, and other items which provide something of the incident's context.

Two factors affect the number of police files initially selected. An allowance was made for non-interviews based on the fact that some individuals were listed in the police files 22 months prior to interviewing. As well, it was known that a portion of the initial selection of police files would be incidents that were out-of-scope for this survey. Excluded as being out-of-scope were cases where the victim was under 18 years of age or where the victim lived outside the city limits of Edmonton at the time the event was reported to the police. Also excluded were incidents involving commercial establishments as the victims.

Details were sent to the Edmonton City Policy Department specifying the number of police files to be selected for each of the subsamples. This sample was stratified by U.C.R. classification and month of occurrence. For each month, the sample of each of the crime classifications was manually selected by the project team and split between the telephone and personal interviewing methods.

The reference sample consisted of victims of the selected crimes which occurred during the period January 1, 1976 to February 28, 1977. This time frame corresponded approximately to the reference period of the study (the reference period was January 1, 1976 to April 30, 1977). A sample of victims of break and enter, robbery, assault, motor vehicle theft, theft and vandalism was selected. In the case of the crime of rape, all eligible cases were included.

Table A: Reference Subsample: Type of Crime by Interviewing Method

Type of Crime	Sample Size Personal	Telephone	Total
Break and Enter	87	88	175
Robbery	100	99	199
Assault	126	124	250
Theft-Motor Vehicle	86	85	171
Theft	84	84	168
Vandalism	88	85	173
Rape	23	23	46
Total	594	588	1182

The telescope sample was chosen from incidents occurring during the period July 1, 1975 to December 31, 1975. This subsample was selected in order to determine the extent to which individuals telescope incidents into the reference period. That this was not the primary concern of the pretest is reflected in the smaller sample size.

Table B: Telescope Subsample: Type of Crime by Interviewing Method

Type of Crime	Sample Size Personal	Telephone	Total
Break and Enter	38	37	75
Robbery	45	45	90
Assault (including rape)	47	46	93
Theft	35	36	71
Total	165	164	329

The unfounded sample was a small group of individuals who reported incidents to the police between January 1, 1976 and February 28, 1977. An incident is called unfounded by the police if it is determined that a crime did not occur. Unfounded crimes range from incidents where an individual reports an automobile stolen and later discovers that it was taken by his son without permission, to a reported rape where, during further investigation, the police determine that no rape actually occurred. These incidents were chosen in order to discover if unfounded incidents would be reported in a victimization survey. The total sample of unfounded incidents was 17 personal and 20 telephone interview cases.

In total, 1,525 individuals were selected from the Edmonton City Police offence records. In the telephone interview sample, 761 individuals were selected, including 11 individuals who were selected for two different incidents. 764 victims were included in the personal interview sample; 12 of these individuals had been selected for two different incidents.

3.2 City Directory Sample

Also incorporated into the design of the Edmonton Reverse Record Check was a sample selected from the 1975 Edmonton City Directory. This city directory sample was included in order to provide a camouflage group for the reverse record check individuals. This approach was based on experience in the U.S. Crime Survey Reverse Record Checks which suggested that a bias may exist if interviewers are aware that each individual in the sample has been the victim of a crime. As well, it ensured that the questionnaire was tested on non-victims.

The shortcomings of a city directory as a sampling frame are well known. However, given the purposes of the city directory sample, a complete and up-to-date frame was not necessary. At the time of interviewing, it appeared that the city directory was approximately 18 months out-of-date.

An advantage of the city directory was that, in conjunction with a telephone book, it provided the same information that was given to the interviewers for the reverse record check individuals. Another advantage was that it listed all individuals in a household 18 years of age and over. In order to achieve as much similarity as possible, an identical age limit was placed on the police sample.

A systematic sample of 1500 individuals was selected from the city directory and assigned to the telephone and personal interview methods. The directory and police samples were chosen several months apart, with the city directory sample being chosen first. Some duplication occurred in the two selection procedures and, since the city directory sample was used primarily as a camouflage group, any individual appearing in both samples was dropped from the directory sample. This resulted in an actual sample size of 1,481 from the city directory - 745 in the telephone interview sample and 736 in the personal interview sample. The sample selection was completed during March 1977 and interviewing was scheduled for the period May 2 to May 20, 1977.

3.3 Operational Procedures

Interviews were conducted both over the telephone and in person. There were 30 interviewers and 3 interviewer supervisors involved in the pretest. The personal interviews were conducted by 20 interviewers with 2 supervisors. Approximately half of each interviewer's assignment consisted of individuals selected from the police records. The remainder was made up of those individuals selected from the city directory.

All telephone interviews were conducted from the Statistics Canada Regional Office in Edmonton. Telephones with headsets were installed to facilitate the interviewing and all interviewers were located in one room separated into cubicles by sound barriers. The interviewer supervisor was located in the same room on a full-time basis throughout the interviewing period.

As some individuals chosen from the police records had been victims as much as 22 months prior to the interview period, a tracing operation was initiated for the police file sample. Since the primary purpose of the city directory sample was to serve as a camouflage for the reverse record check individuals, only the usual interviewer tracing was employed for this sample.

Due to the nature of the survey, it was anticipated that formal channels should be set up for the verification of its authenticity. It was expected that questions concerning the authenticity of the survey would be particularly frequent for the telephone interviewers. Two mechanisms were set up to handle this problem. If respondents became suspicious of the survey, interviewers were instructed to provide them with the telephone numbers of the Edmonton Regional Office and the Complaints Department of the Edmonton City Police. At the Regional Office, all calls were directed to the Regional Office Supervisor for the project. At the police complaints number, all officers likely to be answering telephones were provided with a list of interviewers and a survey summary sheet which provided them with sufficient information to answer common inquiries. As the interviewers were unaware that half of the sample had been selected from police records, all respondents who questioned the source of the sample were referred to the Regional Office where the appropriate information was available.

4. RESULTS

The results are presented under the two broad objectives of the reverse record check study: first, the comparison of telephone and personal interviewing methods, and second, the evaluation of the questionnaire design. The discussion that follows is organized similarly to the statement of objectives in section 2 of this report.

4.1 Response Rate

Due to the sampling procedures employed for the Edmonton pretest, response rates were not expected to compare favourably with conventional Statistics Canada surveys. This was primarily because some individuals were selected for crimes that occurred 22 months prior to interviewing and victims of crime are reported to have high mobility. As well, although the most recent issue of the city directory was employed, it was known to be at least 18 months out of date. The sample size was adjusted to allow for this non-response but, in fact, the actual non-interview rate was higher than expected.

The overall response rates were 57.9% for the personal interviewing method and, for the telephone, 58.8%. The response rate for the sample chosen from the city directory was 64.4% for the personal interview and 65.9% for the telephone. In the police file sample, the response rates were 51.4% and 51.7% for the personal and telephone interviewing methods respectively. These response rates would seem to indicate that there is no crucial difference between the two interviewing methods in this respect.

The following table summarizes the interview status of individuals in the Police File and City Directory samples.

Table C: Final Status by Interviewing Method

	Police File Sample				City Directory Sample			
	Personal		Telephone		Personal		Telephone	
	Number	%	Number	%	Number	%	Number	%
Total Listed	764		761		736		745	
Refusals	10	1.3	39	5.1	13	1.8	40	5.4
Interview prevented by death, sickness, language problem	12	1.6	27	3.5	14	1.9	42	5.6
Temporarily absent, no one home	62	8.1	30	3.9	38	5.2	22	2.9
Moved, abandoned, converted to business	272	35.6	248	32.6	197	26.8	150	20.1
Traced to wrong individual	13	1.7	24	3.1	---		---	
Questionnaire lost in transmission	---		2	0.3	---		---	
Completed	395	51.7	391	51.4	474	64.4	491	65.9

The most interesting comparison in Table C is the refusal rate. The greater number over the telephone may partially be the result of the fact that it is much easier to refuse when the interview is not a face-to-face situation. However, it is felt that this number of refusals does not pose an insurmountable obstacle. The refusal rate is not large enough to warrant the conclusion that telephone interviewing is not feasible. As well, it may be possible to reduce the number of refusals by improving operational procedures.

For the police file sample, the response rate can be calculated by type of crime. For those selected during the reference period, the response rate ranged from 39.7% for assault to 72.7% for vandalism in the personal interviewing method. For the telephone interview, the response rate ranged from 34.2% for robbery to 67.1% for motor vehicle theft. The response rates by type of crime are provided in the table below.

Table D: Response Rate by Type of Crime: Reference Subsample

Type of Crime	Personal Number Responding	Personal %	Telephone Number Responding	Telephone %
Robbery	41	41.0	34	34.3
Assault	50	39.7	50	50.0
Rape	9	39.1	11	47.8
Theft	52	61.9	55	65.5
Theft - Motor Vehicle	55	64.0	57	67.1
Break and Enter	56	64.4	46	52.3
Vandalism	64	72.7	49	57.7

For the individuals selected from the Edmonton City Police files, age and sex were recorded when available. It is therefore possible to determine if the distribution of respondents over these variables is different from the age and sex distributions of the sample for each interviewing method. If respondents are not significantly different from non-respondents, one would expect that the distributions of age and sex for respondents would be similar to those distributions calculated for the entire sample. The χ^2 Goodness of Fit Test indicated that for both interviewing methods, respondents were not significantly different from non-respondents with respect to age or sex. This was true both for the total of all crimes and by crime.

Most surveys encounter the problem of individuals refusing specific questions. As expected, income, age, industry and occupation were the questions most often refused by respondents. As well, the section of the questionnaire on measures used as protection against crime proved particularly sensitive. Some respondents refused to answer questions in this section. It was at this point that many telephone respondents questioned the authenticity of the survey and telephoned either the police or the Regional Office. However, after verifying the survey, most of the respondents continued with the remainder of the interview. The mechanism of survey verification was essential to minimize the problem of respondents refusing to answer the sensitive questions concerning precautions against crime. The verification procedure made completing an interview no more of a problem over the telephone than for personal interviews. Interrupting the continuity of the interview for verification was not sufficiently disruptive to consider dropping these questions due to their importance to the survey objectives.

It should also be noted that for both interviewing methods, refusals between the completion of the screen questionnaire and any crime incident reports required were very infrequent. As well, there were very few refusals to particular questions within the crime incident report. As expected, for both interviewing methods, the problem is not gathering specific details of an incident once it has been reported but eliciting the initial mention of the incident. Once respondents report an incident, they are eager to talk about the details of the crime. This was true even for those reporting multiple incidents.

4.2 Hit Rate

The most important criterion for determining the feasibility of conducting a victimization survey was whether the incidents selected from the police file were reported during the subsequent interview. This was determined by 'matching' details recorded from the victim's police file with the information obtained in the interview. Based on the

experience and recommendations of those involved in the U.S. reverse record checks, the match status was decided on a consensus basis by members of the project team. This process could not be prescribed by a series of rules but was a subjective process with each document being judged independently based on the total information included in the survey document. The summary of each incident, recorded at the end of each interview, was especially helpful in this procedure. It should be noted that the forms were blinded to ensure that the project team was unaware of whether the report was from the personal or telephone interview survey.

The hit rate represents the proportion of individuals interviewed who reported the crime for which they were selected. The overall hit rate was 64.3% for the personal interview sample and 62.7% for the telephone sample. This comparison supports the possibility of conducting a victimization survey by telephone.

Table E: Hit Rate by Interviewing Method

	Personal		Telephone	
Type of Crime	Number Reporting	%	Number Reporting	%
All Crime	207	64.3	195	62.7
Robbery	26	65.0	24	72.7
Assault	33	67.4	32	51.6
Rape	6	66.7	6	60.0
Theft	34	65.4	37	67.3
Theft - Motor Vehicle	38	69.1	43	75.4
Break and Enter	39	70.9	33	73.3
Vandalism	31	50.0	20	40.8

Of the seven crimes under study in the Edmonton Reverse Record Check, rape and vandalism were considered to be the most problematic. Rape represents a distinct problem due to the sensitivity of the subject. It is unrealistic to abruptly ask a respondent "Were you raped during 1977?". However, because of its seriousness and the fact that it is often unreported to the police, it is very important to include the crime of rape in the survey. The approach of this survey is to define an attack, for the respondent, as including 'anything from being hit, slapped or pushed to being shot, raped or beaten up'. It was hoped that the screening items which followed, one of which included the phrase 'attack or molest', might elicit a report of rape, if one had occurred. This approach was a compromise between the importance of including the crime and the possible sensitivity to the respondent of questions relating to it. It should be noted that the hit rates for rape are based on a very small number of incidents. This was a function of both the mobility of the victims as well as the small sample size resulting from the number of such crimes reported to the police.

Table E indicates that the other forms of assaults were also a problem. It was hypothesized that a factor contributing to this might be the relationship between the victim and the offender. The comparison indicated that respondents were more likely to report incidents involving strangers than non-strangers.

Table E also shows that vandalism represents the lowest hit rate for both interviewing methods. Prior to the survey, there was no evidence indicating whether such incidents might be too trivial for respondents to recall since most other victimization surveys have not included the crime of vandalism. However, it was considered important to include it in the survey since it was felt that being a victim of vandalism might affect an individual's attitudes or behaviour. It was hypothesized that the low hit rate for vandalism might be explained by the value of the damage which occurred. This seems to have been borne out in the personal

sample; however, it is difficult to draw conclusions concerning the effect of the value of damage on the hit rate due to the small number of cases reported in the survey.

If the overall hit rate is calculated excluding vandalism, it is 67.8% for personal interviews and 66.5% for telephone interviews. The results of the reverse record checks conducted in the U.S. were 81% in Washington, D.C., 67% in Baltimore, and 74% in San Jose. The results of the Edmonton Reverse Record Check compare with the American results when it is considered that the U.S. studies involved reference periods which were significantly shorter. Their reference periods ranged from 3 to 12 months whereas the Edmonton reference period was 16 months.

As stated earlier, for all individuals selected from the police files, age and sex were recorded when available. This made it possible to determine if the distribution of those reporting the crime selected from the police files differed from the age and sex distribution for all respondents. If respondents who reported the incident for which they were selected are not significantly different from those not reporting, one would expect that the distribution of age and sex for those reporting would be similar to the distribution for all respondents. The χ^2 Goodness of Fit Test indicated that, for both interviewing methods, those reporting incidents for which they were selected were not significantly different from those not reporting with respect to age and sex. This was true both for the total of all crimes and by crime.

4.3 Cost

As stated earlier, an important advantage of telephone interviewing is a considerable savings in cost. Included in the cost comparison are the salaries and associated expenses for the interviewers and Regional Office staff for field work in the Edmonton pretest.

It is evident from the following table that the cost per interview for the telephone method was approximately 1/3 as much as that for the personal interviewing method. It is probable that in a large survey the cost per interview would be less than the figures below due to fixed costs being distributed over a larger number of interviews. As stated previously, experience has indicated that, in general, telephone interviewing should be as much as 70% less expensive than personal interviews.

Table F: Cost by Interviewing Method

Expenditures	Personal	Telephone	Total
Interviewer Fees	$ 13,477	$ 5,047	$ 18,524
Interviewer Expenses	2,150	-	2,150
Associated R.O. Costs[1]	3,337	1,668	5,005
Total	18,964	6,715	25,679
Cost per Interview	$ 12.64	$ 4.46	

[1] The Regional Office Costs are split 2/3 for personal interviewing and 1/3 for the telephone interviewing based on the number of interviewers employed in each method.

Due to inexperience with telephone interviewing, the interviewing schedule was not as productive as possible. Future surveys would benefit from the knowledge gained in this pretest and the resulting change in the time schedule may also reduce the cost.

4.4 Telescoping

This section will first examine the effects of internal telescoping, the misplacing of the incident in time within the reference period. Tables G and H provide a comparison of the date as reported during the interview with the date of occurrence as recorded in the police files.

Table G: Internal Telescoping - Telephone Interviewing Method - All Crimes

Reported to Police	Total	Reported in Interview: JAN '76	FEB '76	MAR '76	APR '76	MAY '76	JUNE '76	JULY '76	AUG '76	SEPT '76	OCT '76	NOV '76	DEC '76	JAN '77	FEB '77	MAR '77	APR '77	MONTH N.A.	Not Reported in Interview
TOTAL	195																		117
JAN '76	13	6	2	1										1				3	11
FEB '76	8		5		1				1									1	9
MAR '76	12		1	5	1	2				3									9
APR '76	13			2	3	4	2			1								1	7
MAY '76	14				2	7	2	3											9
JUNE '76	12					2	8			2									14
JULY '76	11						2	5	1	1	2								9
AUG '76	14							4	6		2	1						1	9
SEPT '76	16							1	2	8	2	2	1						5
OCT '76	17								1	5	9		1					1	8
NOV '76	13							4			2	8	2						10
DEC '76	16					1							12	1				2	7
JAN '77	14												1	10	1	2			7
FEB '77	22											1		3	15		1	1	3

Table H: Internal Telescoping - Personal Interviewing Method - All Crimes

Reported to Police	Reported in Interview																	Not Reported in Interview
	Total	JAN '76	FEB '76	MAR '76	APR '76	MAY '76	JUNE '76	JULY '76	AUG '76	SEPT '76	OCT '76	NOV '76	DEC '76	JAN '77	FEB '77	MAR '77	APR '77	MONTH N.A.	
TOTAL	207																		116
JAN '76	12	7	1	2							1			1					14
FEB '76	13		8	1	2						1							1	11
MAR '76	13		2	6	1							2						2	10
APR '76	15			2	5	4	2				1							1	7
MAY '76	14				1	8	2		1		1							1	15
JUNE '76	12		1			3	5	2	1										6
JULY '76	15			1			1	8	4									1	6
AUG '76	15							2	9	2	2								7
SEPT '76	16							2	2	4	5	1						2	9
OCT '76	9							2	1	1	3							2	11
NOV '76	17									1	3	6	3		1			3	5
DEC '76	21										1	2	18						4
JAN '77	18												1	13	4				7
FEB '77	17		1										1		12	2	1		4

From these tables, it is evident that, for the total of all the crimes selected, respondents reported the month of occurrence in a range often centered about the actual date of occurrence as given in the police files. This trend is also apparent when comparing the dates of occurrence by type of crime for each interviewing method. It is interesting to note that crimes which were reported to the police as occurring in September produced the widest range of occurrence dates in the survey for both interviewing methods. Although a 12 month reference period of September to September was considered at one time to be a salient period because of the school year beginning, this in fact may not be true.

Of the incidents in the personal interview sample, 54% of those reported in the survey were reported in the correct month. In the telephone interviewing method, 55% were reported in the correct month. The accuracy of reporting varied by type of crime. For the personal interview, it ranged from 39% for vandalism to 65% for robbery. The telephone reporting ranged from 45% for vandalism to 71% for assault (including rape). Another indication of the accuracy of reporting is the number of individuals who reported the incident within one month of the actual month of occurrence, that is, plus or minus one month of the month of occurrence. For the personal visit sample, this ranged from 58% for vandalism to 90% for break and enter, for an overall rate of 79%. For the telephone method, the range was from 65% for vandalism to 84% for motor vehicle theft, with an overall rate of 76%.

It is apparent from the tables that the shorter the time interval between the incident and the interview, the better is the accuracy of reporting the date of occurrence. It should also be noted that for both interviewing methods, there was a slight net forward telescoping effect. That is, respondents were more likely to telescope an incident forward within the reference period than to report it as occurring prior to the date reported in the police files.

The second type of telescoping investigated in this pretest is forward telescoping, that is, the reporting of an incident which occurred prior to the reference period as occurring within the reference period. The following table provides both the number and percentage of those incidents incorrectly reported as occurring within the reference period when they had in fact occurred in 1975. It should be noted that if there was no forward telescoping, the hit rate for these cases would be zero.

Table 1: Hit Rate by Type of Crime: Telescope Subsample

	Personal		Telephone	
Type of Crime	Number	%	Number	%
All Crime	13	19.1	17	24.6
Theft	4	23.5	7	31.2
Assault (including rape)	1	6.3	1	7.7
Break and Enter	5	21.7	4	21.1
Robbery	3	25.0	5	33.3

Of those persons selected from the police files for a crime occurring during 1975, 19% of those interviewed reported the incident in the survey during the personal interview whereas over the telephone, 25% reported the incident as occurring within the reference period. However, this represents a difference of only 4 cases between the two methods. For the personal interview, those reporting the incident ranged from 6% for assault to 25% for robbery. Over the telephone the range was from 8% for assault to 33% for robbery. It would also appear that individuals are more likely to report the incident in the interview if the incident occurred close to the beginning of the reference period.

It should also be noted that of the 30 incidents in either interviewing method, 10 were reported in the correct month but a year later than the actual year of occurrence. It may be possible to reduce this type of forward telescoping by emphasizing the reference period during the interview.

4.5 Unfounded Crimes

An unfounded crime is an incident which is reported to the police but which the police subsequently decide did not, in fact, involve a crime. The purpose of selecting unfounded cases was to investigate their effect on future surveys. However, the completed interviews for the unfounded sample were very few (only 10 cases). No conclusions could be drawn about the effect of unfounded crimes from this small number of cases.

4.6 Evaluation of Questionnaire Design

The information used to evaluate the effectiveness of the questionnaire was obtained during a thorough debriefing of the interviewers and Regional Office staff as well as from a number of interviews conducted by members of the project team. The discussion which follows outlines only the problems encountered concerning the format and concepts of the questionnaire. It should be noted that none of the problems were of a magnitude which interfered with the ability of the interviewer to complete an interview. The questionnaire and the basic concepts underlying its design proved to be very successful. Most of the problems encountered may be rectified without much difficulty.

The questions in the screen questionniare which probe for incidents of crime are perhaps the most crucial in the two documents. We hoped to discover in the pretest which of the probes were effective in eliciting reports of incidents. Those that did not elicit reports could perhaps be dropped to relieve some respondent burden or else be replaced by probes that might be more effective. It became evident during the debriefing, however, that the interviewers had edited their document in such a way that there was no reliable information concerning the productivity of specific items. The interviewers stated that in many cases they had recorded the report of a crime in the screening question which they thought should have elicited a response.

The debriefing did, however, reveal some important problems concerning the probes. Both personal and telephone interviewers found certain sections repetitive due to their order and to the variations in meaning of the questions which were not always communicated to the respondents. However, this did not seem to prompt refusals by the respondents. An effort was made to rectify this by dropping some questions and re-ordering and rewording some others.

The section of the questionnaire that presented the most problem for interviewers was the questions concerning precautionary measures such as using burglar alarms, locking doors, or keeping lights on. It appeared that it was at this point during an interview that many respondents questioned the authenticity of the survey and sought some verification that, in fact, the interviewer was from Statistics Canada. This was also the most likely place in the document for respondents to refuse individual questions. This seemed to be a problem almost exclusively for the telephone interviewers. The personal interviewers were equipped with identification cards which were shown to respondents prior to conducting the interview.

The only other section of the questionnaire which seemed to present problems was that concerning the demographic characteristics of respondents. Income, age, and labour force status were problematic for both telephone and personal interviewers. Again, there was a greater problem over the telephone. A contributing factor to the greater sensitivity of particular questions over the telephone was that there seemed to be a tendency for telephone interviewers to be affected more by an individual respondent's bad reaction. The interviewers then had, for subsequent interviews, a preconceived notion that particular questions would engender a bad reaction and therefore, through their own phrasing of the question, contributed to it.

Respondents who have been the victim of a crime a number of times during the reference period pose a particular problem for victimization surveys. The individuals fall into two broad categories. For some, the incidents occur with such frequency that they characterize their lifestyle rather than being salient events, for example, a security guard in a shopping plaza who is threatened daily or a person who is beaten by his or her spouse frequently. In the other category are individuals who have been the victims of many very minor incidents and who may be unable to distinguish among them.

Both these types of incidents present a similar problem since an individual may not be able to distinguish the details of all the incidents well enough to report them separately. Therefore, a definition must be developed for the interviewers to enable them to gather information concerning the group of incidents. This situation is referred to as a 'series'.

In the Edmonton pretest, a group of incidents was defined as a series if it fulfilled the following requirements:

1. the details of the incidents must be similar;
2. there must be at least three incidents in a series;
3. the respondent must be unable to recall details of the individual incidents well enough to report them separately.

It is apparent that some judgement is required of the interviewer in order to determine if a group of incidents should be defined as a series or if separate incident reports can be completed.

It appeared that the interviewer training did not emphasize this definition as strictly as was necessary. Each interviewer interpreted the definition differently and most of them misinterpreted its intent. For example,

some interviewers defined any three crimes reported by an individual respondent as a series although they may have been very different events.

It should be noted that the problem of series also has implications concerning the analysis of the data, particularly on the estimates of the frequency of crimes. There will be further discussion of the problem of series in the following section.

As previously mentioned, there were very few refusals to continue with a crime incident report once the screen questionnaire had been completed. Lengthy interviews were conducted as a matter of course both over the telephone and in the personal interview situation. Interviews generally lasted 15-20 minutes for the screen questionnaire and 5-10 minutes for each crime incident report completed. Interviews of up to 1 1/2 hours duration involving as many as 10 crime incident reports were conducted with relative ease in both interviewing methods. This was certainly due, in part, to the subject matter of the survey. It was also particularly important for telephone interviewers to be flexible in completing the interviews.

4.7 Classification

One of the advantages of a victimization survey is the flexibility for classifying and describing incidents. It is possible to approximate the U.C.R. classification by counting the most serious crime within the incident and assigning a single crime code or to describe the various elements which constitute the incident. The strength of the survey lies in the more complete description.

A single crime code can be assigned based on the answers to the items in the crime incident report. For example, to be classified as a robbery certain conditions must be present during the incident. First, there must have been something taken or an attempt made to take something and

second, there must have been a weapon present or an attack or threat of attack on the respondent. The classification is hierarchical with each incident being assigned a code based on criteria similar to the U.C.R. classification.

The results of the Edmonton pretest were encouraging in that the information for classification was relatively simple to collect. There are some weaknesses to consider, however. For the crime of rape, and to a lesser degree the other crimes, the sensitivity of the information makes it necessary to wait for the respondent to volunteer information and it is impossible to collect the detailed information which is available to the police for such incidents.

A comparison of the police and survey classification illustrates some of the difficulties of a hierarchical classification of incidents. The survey classified a number of each of the other types of crime as break and enters. This does not necessarily mean that the incident reports did not contain the information necessary to be classified, for example, as a robbery or rape, but rather as a consequence of break and enters being classified before most other crimes, any incident which involved a break and enter would not be available for secondary classification.

Of particular concern were the large number of motor vehicle thefts which are classified as simple thefts. The main reason for this appears to have been that the answer categories for coding the items stolen were insufficiently clear for interviewers. The category 'other motor vehicles' was intended for the coding of trucks, vans, motorcycles, etc. However, it appears that in many cases the interviewers specified the vehicle in the 'other (specify)' category. This was quite simply rectified by listing the categories on the questionnaire.

Those incidents which, due to insufficient information, cannot be classified are also problematic. Each of these was reported to the police and during the survey as an incident but, due to errors or misinterpretation, key variables necessary for classification were not available.

The comparison of survey and police classifications of incidents illustrates some of the problems inherent in assigning a single crime code to each incident and emphasizes the advantages of event analysis.

5. SUMMARY

The Edmonton Reverse Record Check was the first stage in the development of a Canadian Victimization Survey. Due to the necessity of developing a cost-efficient methodology, the crucial concern was the comparison of personal and telephone interviewing techniques. As well, the pretest provided information relevant to the questionnaire design and operational procedures. The following discussion will deal with the major decisions made as a result of the pretest.

Conventional wisdom with regard to telephone interviewing has been that long interviews are not possible. If this were true in all cases, it would preclude the possibility of conducting victimization surveys by telephone due to the necessity of collecting information on multiple victimizations. However, it was discovered during the Edmonton pretest that interviews of an hour or more were conducted with relative ease. Victimization surveys would appear to be more conducive to telephone interviewing than some other subject matters due to the interest of respondents in co-operating in a survey which may contribute to reducing crime rates. It is fortunate that those individuals who are subjected to the longest interviews are often the very respondents most interested in completing the survey because they have been victimized the greatest number of times.

The hit rate and response rate were other indicators which contributed to the final decision on telephone interviewing. The results show that for both these factors there were not significant differences between the telephone and personal interview techniques. Given these findings, it would appear that telephone interviewing is the preferable data collection method considering the savings in cost.

The question of the effect of telescoping on a victimization survey has not been fully clarified. The degree of accuracy of reporting the date of occurrence within the reference period is acceptable based on the objectives. However, due to the underestimating of the mobility of the victims selected, the problem of forward telescoping remains unanswered. It would not be warranted to base any conclusion on the small number of cases that constituted this sample. The sample of unfounded incidents was also very small and it was difficult to assess their impact on a survey. A victimization survey asks the respondents to report any incident which they feel has involved a crime of the type being surveyed. If, at the time of the interview, the respondent considers that he/she has been a victim and reports it with sufficient detail to classify it, then it would be included in the survey. The interviewer does not investigate the event in the manner of the police but only records the details of the incident. The concept of unfounded crimes, although relevant, may not have a significant impact on a survey due to the infrequency of such events.

During the course of the field work of the Edmonton Reverse Record Check, a considerable amount was learned about the operational problems involved in collecting victimization data. The importance of interviewer training to the outcome of the survey can not be stressed too much. Two full days of training were used and this proved to be insufficient for conveying all the procedures and concepts involved in such a complex survey. Improvements in training should have a positive effect on all aspects of the survey collection.

One of the most perplexing problems in collecting victimization data is the question of series victimizations. The operational aspect of the problem is to reduce the frequency of series reports. The fact that the concept exists is a temptation for interviewers to record any group of incidents which remotely fits the definition into this category in order to reduce the number of incident reports necessary. Making the definition too restrictive risks increasing the respondent burden. It remains to improve the training of interviewers and supervisors in the concept in order to minimize the misuse of the classification. Some change in the definition may make it possible to avoid misuse while reducing the frequency of series reports in some significant way.

The problem of how to count series victimizations when they are reported is the other aspect of the problem. The fact that the respondent cannot remember details of the incidents in sufficient detail to report them separately indicates that the number of incidents may not be accurate. For example, an individual who reports that her husband beats her almost every day and estimates the total as 250 is obviously not giving an exact number. But excluding incidents of this type from the final estimates risks underestimating the number of incidents by a considerable degree. There seems to be no completely satisfactory method of dealing with this problem.

The Edmonton Reverse Record Check answered the most important question by showing that it was feasible to conduct a victimization survey using telephone interviewing. However, inadequacies in the questionnaire showed the need for a number of alterations and, as well, the problem of telescoping was inadequately studied. It was, therefore, decided that a second pretest was necessary. An abbreviated discussion of the second reverse record check conducted in Hamilton is contained in the following sections.

6. HAMILTON REVERSE RECORD CHECK

The second pretest was conducted in Hamilton during February 1978 using only telephone interviewing. Once again, a reverse record check technique was employed to ensure that a sufficient number of victims was surveyed for each type of crime. The crimes sampled were assault, sexual assault, robbery, motor vehicle theft, break and enter, theft and vandalism. The objectives of the Hamilton survey were to test the ability of the questionnaire to elicit incidents, to investigate the effects of telescoping, and to test the revised field procedures.

The results of the Edmonton survey indicated that several revisions could be made that might increase the overall reporting of incidents. Two changes in particular were considered important in improving the success of eliciting incidents. The 16 month time frame employed in Edmonton was considered not only awkward but too lengthy for good recall. The Hamilton survey was to have a 12 month reference period which coincided with the calendar year 1977. The screening items were also revised in an attempt to elicit more incidents, particularly those referring to the crimes with the lowest hit rates in the Edmonton study.

6.1 Study Design

With the co-operation of the Hamilton-Wentworth Regional Police Force, 2,862 individuals listed as victims of one of the seven crimes included in the survey were selected from the offence reports in the police files. In contrast to Edmonton, a camouflage sample was not incorporated into the design of the Hamilton pretest.

In order that the interviewing could begin in January 1978, it was necessary to conduct the actual selection in December. Therefore, only incidents occurring prior to December 1, 1977 were available for selection. The reference sample consisted of incidents of the selected crimes during the period January 1, 1977 to November 30, 1977. This time frame corresponded approximately to the reference period of the survey. A stratified simple random sample of incidents of break and enter, assault,

theft-motor vehicle, theft and vandalism was selected. For the crime categories of robbery and sexual assault, all incidents that were within the scope of the survey were selected. A total of 1,883 incidents was included in this sample.

The telescope sample was chosen from incidents occurring between July 1, 1976 and December 31, 1976. In Edmonton, a very small number of successful interviews was completed with this subsample because non-interview rates had been underestimated. Therefore, a large number of incidents was included in this sample for the Hamilton pretest. As well, in Edmonton, only incidents of break and enter, assault, robbery and theft were included. All eligible cases of robbery and assault and a stratified simple random sample of incidents of break and enter, theft, theft-motor vehicle, assault and vandalism were selected in Hamilton, for a total of 1,048 incidents.

In summary, there was a total of 2,931 incidents selected from the Hamilton-Wentworth Regional Police offence reports. Fifty-seven individuals were selected for two different incidents and six individuals were chosen for three incidents.

6.2 Results

An overall improvement in results was achieved in the Hamilton pretest, indicating that the methodology for conducting telephone victimization surveys was successful. No major problems were unearthed which necessitated further changes in the basic design of the survey.

The hit rate improved significantly indicating that the changes in questionnaire design and reference period were effective although there is no way to determine the degree to which they influenced this improvement. For incidents occurring during the reference period, the hit rate was 71.8%, ranging from 55.7% for assault to 82.1% for robbery. Overall this represented an increase of approximately 9% from the telephone sample of the Edmonton study. At this level, the hit rate would suggest that the methodology is successful and that future surveys are feasible.

Telescoping remains a difficult consideration. Overall, the reporting rate was 21.4% This is similar to the results of the Edmonton pretest. It is possible to drastically reduce the problems posed by telescoping; however, the solutions may not be congruent with the constraints of Canadian victimization surveys. Through bounding interviews in a panel design, the difficulty of forward telescoping can be almost entirely removed and shortening the reference period would reduce internal telescoping. These alternatives could be employed if an ongoing survey was anticipated but this is not the case.

The amount of internal telescoping suggests that, given a year reference period, estimates for shorter periods may not be reliable. Again the effect of this can be minimized if it is remembered that the emphasis is evaluation and therefore it is the change in estimates that is crucial. It would be assumed that any telescoping would be constant over two or more unbounded surveys.

The definition of a series was altered for the Hamilton pretest. The number of incidents required to constitute a series was increased from three to four. The attempt to reduce the number of series incidents reported was successful to a limited degree. However, it remains that certain individuals are victimized a number of times over a year and cannot report the incidents as distinct events. There are, in fact, several related theoretical and operational problems that must be considered.

When estimating the incidence of crime, reports of series create a problem since it has not yet been determined how to count series reports. The very fact that they are dealt with differently is an admission that they are exceptions. The analysis is complicated by the fact that the individual's estimate of the number in the series is often a 'wild guess'. Equally important is the fact that only one incident form is completed for a series of incidents and although all the events in a series are similar, according to the definition, a summary description of the details of an event perhaps cannot be compared with the details

of the victimization reported by others in the survey. The analysis of risk and impact would be confounded by this situation. In this discussion no solutions have been suggested other than urging that every effort be made to reduce the number of series events reported by reducing those which involve interviewer errors. Series incidents will not disappear and although including them in some analysis such as comparing the demographics of victims and non-victims in general terms may be acceptable, caution should be employed and it is perhaps advisable to deal with series events separately.

As in Edmonton, it was possible to compare the police classification with the survey classification. There was a considerable overall increase in the number of incidents classified identically by the police and by the survey. A number of factors were responsible for this, most importantly the improvements in the questionnaire and the training of interviewers. The actual classification rules remained basically the same. Approximately 80% of the incidents had the same classification under both schemes.

The percentage of incidents which the survey could not classify was reduced in Hamilton. It is necessary to emphasize that while the comparison of police and survey classification is valuable for identifying problems in questionnaire design, the more descriptive classifications which are possible will be more useful in analyzing the event in future surveys.

During the Edmonton and Hamilton pretests, it became very obvious that an essential procedure for a successful victimization telephone survey was the verification of the authenticity of the survey by the local police department. Many respondents became wary when the questions concerning burglar alarms, locked doors and activity patterns were asked and certainly this was with some justification. Respondents were given both the local Statistics Canada and police department telephone numbers if they questioned the survey. This happened with regularity and, although no estimate as to the exact number was available from the police, it was

obvious that hundreds of interviews would not have been completed without this procedure. The co-operation of the Edmonton City Policy Department and the Hamilton-Wentworth Regional Police Department was essential to the success of the pretests during both the sample selection and the interview periods.

7. CONCLUSIONS

The overall objective of the pretests reported here was to produce a cost-efficient victimization survey methodology which would address the needs of researchers concerned with crime and its effect on society and those directly involved in efforts to reduce crime. In order to identify and include within the questionnaire these information requirements, consultation took place with representatives of law enforcement agencies, researchers and those experienced with victimization studies in the U.S.

The methodological tests revealed that telephone interviewing was feasible. It was felt that an important factor in the success of the telephone was the police verification procedure. Personal interviewers, equipped with Statistics Canada I.D. cards, do not contend with respondents questioning their legitimacy, whereas this is a considerable problem for the telephone interviewers. The subject matter of the survey tends to reinforce the uncertainty of some respondents contacted by telephone.

The following guidelines were formed as the result of the two pretests and would be recommended for future surveys:

(1) the police number should be one listed in the telephone directory;
(2) all calls should be handled consistently at a centralized location within the police department;
(3) calls must be answered at all hours of the day;
(4) the police should not be expected to justify the survey but strictly verify its authenticity.

An obvious omission from this discussion is reference to a sampling frame for future surveys. Two avenues which have been investigated are the use of a random digit dialing technique and telephone company listings. Random digit dialing has been researched in the U.S. with very satisfactory results. Telephone company listings, if available, may also be an excellent frame. They contain, on computer file, all subscribers who would typically be included in a telephone directory as well as non-published numbers. An advantage of the telephone listings is that it allows for the exclusion of business subscribers, therefore drastically reducing the number of non-productive calls. It is also updated frequently thus avoiding a number of pit-falls normally associated with employing a telephone directory as a sampling frame.

It will be necessary for future surveys to deal with telephone sampling on an ad hoc basis given the population to be studied, the geographic area to be covered and other idiosyncrasies of a specific survey.

The final stage in the development of a tested and documented Canadian Victimization Survey was a pilot survey producing data related to the original objectives of the research. During January and February 1979, a full-scale survey was conducted in Vancouver. The data resulting from that survey are presently being analyzed.

ACKNOWLEDGEMENT

The authors wish to acknowledge the contribution of John Evans and Gerry Leger of the Research Division of the Department of the Solicitor General during the development and conduct of this research. We would also like to thank the other project team members, Judy Sauvé, Beth Duncan, Rowe Haddow, Laurie Barnes and Ray Ryan.

RESUME

Le présent article expose la méthodologie et l'analyse de deux principaux essais préliminaires afin de comparer l'efficacité de différentes méthodes d'interview et d'évaluer la possibilité de recueillir des données répondant aux exigences de l'enquête sur les victimes d'actes criminels.

REFERENCES

[1] Biderman, A.D., Johnson, L.M., McIntyre, J., and Weir, A.W., Report on a pilot study in the District of Columbia on victimization and attitudes toward law enforcement. Washington, D.C.: U.S. Government Printing Office, 1967.

[2] Catlin, G.P. and Murray, S.M., Report on Canadian Victimization Survey Methodological Pretests, Statistics Canada, 1979 (unpublished).

[3] Eidson Penick, B.K. and Owens, M.E.B. (Eds.). Surveying Crime (Report of the Panel for the Evaluation of Crime Surveys, submitted to the Law Enforcement Assistance Administration). Washington, D.C.: National Academy of Sciences, 1976.

[4] Evans, J. and Leger, G., Canadian Victimization Surveys: A Discussion Paper. Canadian Journal of Criminology, 1979, Vol. 21, No. 2, 166-183.

[5] Feinberg, S.E., Victimization and the National Crime Survey: Problems of Design and Analysis. University of Minnesota: Technical Report No. 291, 1977.

[6] Reiss, A., Measurement of the Nature and Amount of Crime. Washington, D.C.: U.S. Government Printing Office, 1967.

[7] Skogan, W.G., The validity of official crime statistics: An empirical investigation, Social Science Quarterly, 1974, 55, 25-38.

[8] Skogan, W.G., The use of victimization surveys in criminal justice planning. In L. Oberlander (Ed.), Quantitative tools for criminal justice planning. Washington, D.C.: Law Enforcement Assistance Administration, 1975.

[9] Skogan, W.G., Measurement problems in official and survey crime rates. Journal of Criminal Justice, 1975, 3, 17-32.

[10] Skogan, W.G. (Ed.). Sample surveys of the victims of crime. Cambridge, Massachusetts: Bellinger, 1976.

[11] Tuchfarber, A.J., and Klecka, W.R., Measuring crime victimization: An Efficient Method. Washington, D.C.: The Police Foundation, 1976.

[12] Tuchfarber, A.J. and Klecka, W.R., Bardes B.A., and Oldendick, R.W., Reducing the cost of victim surveys. In W.G. Skogan (Ed.), Sample surveys of the victims of crime. Cambridge, Massachusetts; Bellinger, 1976.

A PERSONAL VIEW OF HOT DECK IMPUTATION PROCEDURES[1]

Innis G. Sande[2]

A Hot Deck imputation procedure is defined to be one where an incomplete response is completed by using values from one or more other records on the same file and the choice of these records varies with the record requiring imputation.

General approaches to Hot Deck imputation are outlined, with emphasis on the interaction between the edit constraints and the imputation procedures. Distance functions can be constructed on a mixture of categorical and numeric fields, can be modified to take account of the relative importance of fields and can discriminate against less desirable donors. Matching fields may be correlated with missing fields, may be linked with missing fields by edits or may be natural stratification variables; but increasing the number of matching fields does not necessarily result in a better match. It is important to audit the imputation process and to summarize its performance.

Hot Deck procedures should be evaluated to study the bias and reliability of the estimates, donor usage and frequency of imputation failure in terms of a variety of conditions of the data and variations of the imputation procedure. It appears that the only generally available approach to evaluation is by simulation.

1. INTRODUCTION

There is a growing awareness of the use of imputation in the preparation of data files. As the files get larger, the need for automatic imputation becomes essential.

[1] Adapted from a paper presented to the Symposium on Incomplete Data, Washington, D.C., August 1979.

[2] Innis G. Sande, Business Survey Methods Division, Statistics Canada.

Methods of imputation vary considerably, ranging from the use of default values to the development of complex models. One class of imputation procedures is the so-called Hot Deck type, in which an incomplete response is completed by using values from one or more other records on the same file and the choice of these records varies with the record requiring imputation.

This paper describes the author's perception of Hot Deck procedures as a solution to the imputation problem. This necessitates first discussing her perception of the imputation problem, since a different viewpoint could very well result in a different assessment.

2. GENERAL OBSERVATIONS ABOUT IMPUTATION

Imputation is the process of estimating individual values in a data set. It is a direct generalization of the "missing observation" problem in Analysis of Variance and the "Incomplete Data" problem in Multivariate Analysis. Solutions of these two problems typically make use of very specific model assumptions about the data.

The need for imputation arises in two ways:

(i) a record (multivariate observation for a single case) contains one or more missing values because the data is unavailable; or

(ii) a record is inconsistent, i.e. its values do not satisfy natural or reasonable constraints (edits) and one or more values are designated for change (and are, therefore, artificially "missing").

One may reasonably ask: why impute at all? Would it not be preferable to leave the data incomplete and analyze what remains, tabulating the missing values as "unknown"? Surely, imputation is a process of delusion, giving the impression that the data are in better shape than they actually are.

There is much to be said for this argument. Imputation, by whatever method, can add no new information to the data (except, possibly, when auxiliary data are available). If badly done, it may result in serious misrepresentation of facts. However, there seem to be at least two cases where imputation is useful:

(i) imputation of a very small proportion of values, so small and by such a method that no discernible distortion of the data could result, may make the data set much easier to handle, e.g. the imputation of a few points in a time series of equally spaced observations; or

(ii) imputation where the end products are tabulations at arbitrary levels of aggregation.

Case (ii) is the one familiar to survey takers in particular. Including "unknowns" in all tables is usually thought to be untidy and deleting these cases produces inconsistent tables (the totals or marginal distributions vary). In addition, where a record consists of a fairly large number of fields, one has the feeling that some information about the missing or questionable values is contained in the portion that is good.

It is not the function of this paper to discuss all methods of imputation; but we note some of the common elements of imputation procedures.

(a) There is a close relationship between editing and imputation. If a record fails an edit, it is not always obvious which fields are faulty; yet some basis for deciding which fields to change must be established. (We are assuming that all possible cleaning up, by means of reference to original records or respondents, has been done). Complex edits make life exceedingly hard, in deciding both which fields to impute and how to impute missing or questionable fields (see [5]). This problem is frequently ignored in theoretical work on missing data.

(b) The marginal and joint distributions of responses are almost certainly different from those of the underlying population. In the case of numeric data, such distributions are unlikely to be normal. Moreover, transformations to normality (or just to less pronounced skewness) usually result in transformations of the edits which make them more difficult to deal with.

(c) The pattern of missing fields varies from record to record. In an n field record (excluding the identifiers) there are 2^n possible patterns of fields to impute. The edit procedures may reduce this number of possibilities in practice, if only to simplify imputation.

(d) The imputer does not have a great deal of time to fiddle with the data after they come in. In fact, he often has tight deadlines. He may have little, if any, test data to work on before the data collection begins.

(e) Imputation, by any method, does not solve any specific estimation problem more satisfactorily than the usual analytical estimation techniques. That is, in order to estimate a particular quantity, θ, one can only use such

data as there are and a relevant model. However, imputation does "solve" the problem of being able to produce, very easily and in a consistent way, estimates of any population parameters (arbitrary totals, means, proportions, etc.), even those the survey was not designed to estimate, although possibly with no guaranteed precision.

What the imputer wants, therefore, is a procedure which

(i) will impute plausibly and consistently provided only that the non-missing data satisfied the edits;

(ii) will preserve the underlying distributions in the data, or at least reduce the bias in the responses, and preserve the relationships between fields as far as possible;

(iii) will work for (almost) any pattern of missing fields; and

(iv) can be set up ahead of time.

Particular techniques of imputation may vary in their dependence on particular models and their ability to stabilize estimates, reduce bias (relative to standard estimation techniques) or preserve the relationships between the variables.

3. HOT DECK PROCEDURES

We will define a Hot Deck imputation procedure to be one where an incomplete response is completed by using values from one or more other records on the same file (i.e. from the same survey) and the choice of these records varies with the record requiring imputation.

Thus, simply inserting the stratum mean of the good data in the missing field is not a Hot Deck procedure because the choice of the records used in the mean is independent of the record requiring imputation. All imputations of the same field in that stratum would be the same. Choosing a "good" record, (the donor) which resembles the "bad" record (the recipient) and using the donor to supply the values of the fields missing in the recipient, is a Hot Deck procedure.

As a simple example of a Hot Deck procedure, consider the case of a record consisting entirely of categorical data. An incomplete record requiring imputation in one or more fields is matched to a collection of complete records in the file (the Hot Deck) which have identical values in the remaining fields. One of these complete records is chosen at random and is used to donate the values of the missing fields to the incomplete record.

To formalize the above description somewhat, suppose the n fields of the record are $X_1, \ldots, X_n$. The recipient record lacks $X_1, \ldots, X_l$, but has values for $X_{l+1}, \ldots, X_n$. The recipient record before imputation will then be $\underset{\sim}{x}_R = (,,, x_{l+1}, \ldots, x_n)$, where the blanks stand for unknown values. Now a collection, $C(\underset{\sim}{x}_R)$, of complete records of the form

$$(X_1, \ldots, X_l, x_{l+1}, \ldots, x_n)$$

is identified and one is chosen at random, say

$$\underset{\sim}{x}_D = (x_1', \ldots, x_l', x_{l+1}, \ldots, x_n).$$

This is the donor record. The completed recipient is then $\hat{\underset{\sim}{x}}_R = \underset{\sim}{x}_D$.

It is easy to see that such a procedure would produce consistent imputations, would tend to preserve underlying distributions and to reduce response bias in fields where response is relatively poor. It would work in any situation and could be set up in advance.

In reality, the operation does not work quite as smoothly because (i) there are computational problems, and (ii) there are no exact matches in some cases.

The computational problems are mainly ones of sheer size: for a record with a moderate number of fields, the number of possible Hot Decks which would have to be identified is very large. In practice, compromises have to be made in order to reduce the potential number of the decks. This is usually done by matching on fewer fields and/or imputing for one or a group of fields at a time. This means that a particularly scanty record may receive data from several donors, and also that successive imputations may result in a record which fails one or more edits. To avoid the latter situation, various ad hoc procedures may be employed along the way.

There seem to be two main types of Hot Deck for categorical data which we will call <u>sequential</u> and <u>random choice</u>.

In the sequential procedure (used by the U.S. Census of Population and Current Population Survey, [9], [10], [12], [14], the data are processed one record at a time. A field A (or group of fields) is imputed by defining a cross-classification of several other related fields (B, C, D,..) on which a match is to be made. For each cell in this classification, that value of A is retained which occurred in the last record processed with the corresponding values of B, C, D,... Thus, as the file is processed, the values in the individual cells of the (B, C, D,..) matrix change. When a record which lacks a value of A occurs, it receives the value currently in the cell of the matrix

which matches its own values of B, C, D,... If two such records (lacking a value of A, but with the same values of B, C, D, ...) occur consecutively, the same value of A will be imputed in each case, since no records will have been processed which could cause the value to change. The matching fields (and therefore the imputation matrix) vary with the fields to be imputed. In those cases where imputation of a single field might result in an edit failure after imputation, a set of related fields is deleted and imputed together. One other obvious problem with the sequential procedure is that each imputation matrix must be initialized.

In the random choice procedure, a single current donor matrix is not maintained, but a record is chosen at random from a deck with suitable characteristics. The choice of matching fields in both sequential and random choice procedures must be made considering likely major sources of variation and the number of eligible records available in each cell for donation. We return to the problem of matching in the next section.

In the random choice procedure used by the Canadian Census ([7]), matching is done on those fields linked to the missing fields by edit constraints, as well as fields correlated with the missing fields. This could result in a very large number of matching fields; but those fields are eliminated which do not restrict the value of the field to be imputed given the values of the data present. The procedure first attempts to impute all missing fields using a single donor. If this fails, a field-by-field Hot Deck imputation is tried.

These Hot Deck procedures involve mainly categorical data. It is easy to see that if a single numeric (quantitative) field or a whole group of such fields has to be imputed by matching on categorical fields, these systems will still work. Trouble starts when there are several quantitative fields linked by edit constraints, or matching has to be done on quantitative fields, or both. The matching problem

can occasionally be dealt with by splitting the range of the variable (e.g. age) into intervals and coding the intervals; but if several variables are involved, one may find that the data are unevenly distributed through the grid.

It appears that numeric Hot Decks are not as common as categorical. In [9], a sequential procedure is described for imputing income data; but note that all income fields are imputed even if only some of them are missing. The reluctance o use numeric Hot Decks seems to be due in large part to the difficulties of coping with the edit structure. Furthermore, for certain single-field Hot Decks, it is known that estimates based on imputed data are more variable than those based on weight-adjusted data ([1]). This is because the Hot Deck contains extreme-value as well as central-value records. Funny records are bad enough when they are real - they are no joke when they are imputed.

Conceptually, a numeric Hot Deck requires a distance function to be defined between records on the matching fields since an exact match in numeric fields is unlikely. This function need not be a metric - it need not even be symmetric in the recipient and donor records.

The Hot Deck consists of all "good" (i.e. complete in all the relevant fields) records. For a particular recipient, a donor (or "good") record in its neighbourhood is identified and the missing fields of the recipient record are supplied by transformation of the corresponding fields of the donor. In one implementation at Statistics Canada [11], the nearest m complete records to a particular recipient are identified. This requires an efficient search algorithm. An attempt to complete the deficient record using fields from one of its m neighbours is made, taking the complete records in order of nearness. The donation is successful when the completed record passes the edits. If none of the m neighbours will do the job, the imputation fails and further processing is required.

When implementing this type of system, it is advisable to consider judicious transformations of the data both for matching and for imputation (the transformations appropriate to each function need not be the same). The distributions of some numeric data become very attenuated in the tails, so that "nearness" in the untransformed data changes in different regions. It is also sometimes possible to transform the data in such a way which both conforms with the edits and facilitates a correct imputation. For example, if an edit is $A + B + C \leq E$, then division by E transforms the edit to $P_A + P_B + P_C \leq 1$ and instead of A, B, C, E as data, we have P_A, P_B, P_C, E as data. The distance function may now be defined on some transformation of P_A, P_B, P_C, E.

In one implementation of a mixed numeric and categorical Hot Deck, one nearest neighbour was identified and a set of estimates of the missing fields was specified depending on the values of fields in both recipient and donor records so as to force a consistent imputation. Thus, for example, if the field A was missing in the recipient, the imputation might be

$$\hat{A}_R = f(A_D, B_R, B_D)$$

where B is a field or set of fields present in both recipient and donor and the subscripts D and R signify donor and recipient fields respectively ([2]).

In order to reduce the variability of the numeric Hot Deck imputation, the device of averaging over neighbours (or successive records in a sequential system) has been suggested and used. This will work for single numeric field imputation and will stabilize the final estimates. However, where several numeric fields are being imputed, an averaged record will not necessarily satisfy the edits. In general, mixed numeric and categorical procedures, there is no way to average categorical data.

In some cases auxiliary variables are present for all data. If they are categorical variables only, they may simplify the Hot Deck procedure by guaranteeing minimal matching. If they include numerical variables which are correlated with the survey fields subject to imputation, they can be effectively used as the total basis for matching, making the search procedure much simpler. In such cases, one may argue that it would be better to use a ratio estimate rather than impute missing data; but ratio estimates (like weight-adjusted estimates) are not additive and it appears ([2]) that a suitable imputation procedure could be less biased than the ratio estimate while (more or less) preserving the variance.

For large scale imputation (imputation of large numbers of entire survey records) good auxiliary variables, possibly from administrative sources, are essential and the process can be thought of as transforming auxiliary data into survey data (e.g. [2]).

4. SPECIAL PROBLEMS

4.1 Distance Functions

One of the myths about numeric Hot Decks seems to be that choice of the distance function is critical. In fact, judging from the experience with experimental systems at Statistics Canada, the performance of the Hot Deck is not particularly sensitive to the form of the distance function, once the variables have been transformed and rescaled. However, some distance functions are easier to deal with than others, a particularly attractive one being, after transforming to uniform marginals:

$$d^N(i,j) = \sup_k |x_{ik} - x_{jk}|$$

where i and j index the records. If one of the variables is more important than another, one can incorporate this by weighting them

$$d^N (i,j) = \text{Sup } w_k \; |x_{ik} - x_{jk}| \; .$$

Categorical data can be incorporated by defining suitable resemblance functions between the classes of a categorical variable. For example, if variable A takes values $A_1, \ldots, A_K$, then

$$R(A_k, A_k) = 0$$

and $$R(A_k, A_\ell) = 1 \quad \text{if } A_k \text{ and } A_\ell \text{ are compatible,}$$

$$= 10^5 \text{ if they are not.}$$

One can now define, where A_i is the value of A taken by the ith record,

$$d^C (i,j) = \underset{A}{\text{Sup}} \; (A_i, A_j).$$

The numeric and categorical distance functions can then be combined, e.g.

$$D (i,j) = d^N (i,j) \; .(1 + d^C (i,j)).$$

Obviously, there are many ways to play this game.

If one has reservations about the jth observation, one can inflate any distance which incorporates it, and so render it less preferable than other nearby observations:

$$d' (i,j) = d (i,j)(1 + h_j) \; , \quad \text{where}$$

d (i,j) is any measure of a distance and h_k is presumably zero for most observations k. In particular, in a matching or random choice situation, when i is the recipient and j the donor record, h_j may be a function of the number of times j has already been used as a donor. This has the effect of spreading the donor usage around and avoiding the over-use of a particular donor. Whether this is an advantageous procedure is open to question. If response is poor in some region (so that donors are rare), does one necessarily want to impute using donors in a nearby region where the response is good, but the characteristics of the response may be different? Repeated use of a particular donor will inflate the variance; but equalizing donor usage may result in bias. The main reason for limiting donor usage may be the pacification of nervous clients.

4.2 Choosing the Matching Fields

When a record fails an edit which involves several fields, it is not always obvious which fields are in error. If several edits involving common fields are failed, there are some intuitive grounds for casting suspicion on one or more of the common fields. Depending on the circumstances, one may believe that certain fields are more prone to error than others. The decision about which fields to impute is an editing decision which has little to do with the method of imputation, except insofar as it facilitates the imputation, and we will not deal with it here.

The question we do address is: given that the decision has already been made as to which fields are missing (to be imputed), which of the remaining fields are used for matching? The natural candidates seem to be (i) fields correlated with missing fields, to ensure a good imputation, (ii) fields linked by edits to the missing fields, to avoid edit failure after imputation, and (iii) natural stratification variables employed in the survey design, which may influence the missing data as in (i).

In the case of a record with many variables (such as the Census of Population), the collection of all reasonable fields may be so large that implementation is difficult and there is no guarantee of a match. In the case of a mixed categorical and numeric match, an exact match on the categorical variables may force a poor match on the numeric variables.

Increasing the number of matching variables may not result in a better match. One should give some hard thought to what compromises are acceptable in terms of grouping classes (so that, for example, a recipient in industry I may be imputed from a donor in a compatible industry J) and eliminating variables so that a donor pool of suitable size is available.

A closely related observation is that it is often not possible or even desirable to do all imputation in a single pass (so that each recipient requires only one donor). The number of complete records (potential donors) may be relatively few so that matches would be poor, no use would be made of information in partially complete records and the same donors could be used repeatedly. The matching variables and distance functions appropriate for imputing some variables may not be suitable for imputing others. The imputation is therefore broken up into several stages, with certain sets of fields being imputed at each stage. Different records would be available as potential donors at each stage since they would only be required to be complete in the current matching and imputation fields. A result of this approach is that several donors may be involved in completing a deficient record. On the other hand, imputed fields can be used in matching and donation in succeeding stages.

4.3 Auditing

Some effort should be made to keep track of what the imputation process is doing. At the end of the process, one would like to know:

a) How many times a particular record has been used as a donor in a particular stage.

b) How many attempts had been made to achieve a successful imputation for a particular deficient record (this would not apply to some procedures).

c) Which donors contributed what fields to which recipients. This is important in tracing the sources of peculiar imputations. By analyzing the transfer of information from particular donors to specific recipients, one may trace and remedy problems in the imputation procedures. Remedies may consist of changing the matching variables, the method of estimating missing fields or the definition of a possible donor by excluding those which appear to be outliers although they might be acceptable records.

d) If the imputation of a field is conditional on the values of other fields (in either the recipient or the donor) which condition prevailed at the time of imputation.

e) The value of the distance function at each donation. A relatively large value could signal a problem.

Useful summaries of the run are:

i) the number of records eligible as donors,

ii) the number of records requiring imputation,

iii) the number of records eligible neither as donors nor as recipients,

iv) a frequency distribution of the number of times each donor was used over all donors (see (a) above),

v) a frequency distribution of the number of attempts to achieve a successful donation over all recipients (see (b) above),

vi) frequencies of the condition flags (see (d) above),

vii) a listing of all records for which imputation failed, and

viii) a distribution of the value of the distance function (see (e) above).

Distributions should be for records in fairly homogeneous strata.

5. EVALUATION OF HOT DECKS

An imputer with a new and shiny Hot Deck system naturally wants to know how good it is, and so do the users of the data which the Hot Deck produces. Some of the questions which arise are: how are

i) the bias and reliability of the principal estimates,

ii) donor usage, (the distribution of the frequency with which records are used as donors), and

iii) the frequency of imputation failure,

affected by

i) the size of the data set,

ii) the frequency of missing data,

iii) "non-response" bias (where the non-response may be caused by deletion of fields due to edit failure),

iv) the underlying distributions of the data,

v) the choice of matching fields,

vi) the distance function, and

vii) the particular parameters of the imputation procedure?

A little theoretical work in very restricted situations has been done on reliability and bias ([1], [12]). Part of the difficulty in extending theoretical work lies in the edit structures and part in the sources of variation. Given the sample, numerical matching procedures are generally deterministic. Sequential procedures depend on the ordering of the file which is seldom completely random.

It appears then that the only generally available approach to evaluation is by simulation, using either real or artificial data. Real data, presumably culled from the good records of previous surveys, have the advantage of being realistic. On the other hand, fake data, produced by some modelling process, are subject to more manipulation so that one can vary distributions of and relationships between variables. In either case, fields are designated as missing by some random process which can be replicated and the variation over these replications is observed and analyzed ([2], [6]).

In addition, several empirical studies have been carried out comparing Hot Deck and other procedures with respect to estimation and costs ([1], [2], [3], [4], [8]).

6. THE LAST WORD

In this paper we have attempted to outline what we believe to be the general approaches to Hot Deck imputation, with emphasis on the interaction between the edit constraints and the imputation procedure.

As a method of imputation, Hot Deck has some attractive features in comparison with its competitors, not the least of which is that no strong model assumptions need be made in order to estimate the individual values. The Hot Deck procedure can be viewed as a sort of non parametric regression. Although there may be an increase in the variability of some estimates (depending on the Hot Deck methodology), it does appear that there is a reduction in non-response bias due to partial responses or where auxiliary information is available, at least under normal survey conditions.

There are also many problems associated with Hot Deck procedures, mainly involving accommodation of the edit structure or constraints on the data, and we have tried to discuss these (or rather, those we are aware of) in a general way.

We have not attempted any discussion of the implementation of these procedures, because as far as we know, the implementation tends to be tailored to the application and, in any case, we would be well out of our depth in pretending any knowledge.

We know of no example of a "pure" Hot Deck being used on data of any great complexity. Hot Deck systems appear to be used in conjunction with other imputation methodologies (such as Cold Deck) in order to achieve consistency and reasonable efficiency.

No generalized Hot Deck system has been developed. The CANEDIT system ([7]) is an attempt at one for categorical data; but it has limitations. A generalized numerical Hot Deck system is being developed at Statistics Canada, which deals with linear edits only ([11]). Both these systems involve both edit and imputation phases, using the edit phase to decide which fields to impute on the basis that as few fields as possible should be changed. A generalized, integrated numerical and categorical data edit and imputation system is seen as being feasible, although there are formidable mathematical and algorithmic problems involved.

RESUME

La méthode d'imputation dite du 'hot deck' est celle où l'on complète une réponse incomplète avec des données provenant d'un ou de plusieurs autres dossiers du même fichier; le choix de ces dossiers varie selon le dossier devant faire l'objet d'une imputation.

Le document décrit la méthode générale du 'hot deck', en insistant sur l'interaction entre les contraintes de vérification et les procédures d'imputation. À partir d'une combinaison de zones catégoriques et numériques, il est possible de construire des fonctions de distance, de les modifier de manière à tenir compte de l'importance relative des zones et de défavoriser des donneurs peu désirables. Des zones correspondantes peuvent être corrélées avec des zones manquantes, raccordées à des zones manquantes par vérification ou peuvent être des variables naturelles de stratification; cependant, le fait d'augmenter le nombre de zones correspondantes ne donne pas nécessairement un meilleur appariement. Il importe de contrôler l'imputation et de résumer sa performance.

Il faut évaluer la méthode dite du 'hot deck' pour étudier le biais et la fiabilité des estimations, de l'utilisation des donneurs et de la fréquence de l'échec de l'imputation dans diverses conditions des données et la variation de la procédure d'imputation. Il semble que la simulation soit la seule approche d'évaluation qui soit généralement disponible.

REFERENCES

[1] Bailar, J.C. III and Bailar, B.A., "Comparison of Two Procedures for Imputing Missing Survey Values". Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978, pp. 462-467.

[2] Colledge, M.J., Johnson, J.H., Pare, R. and Sande, I.G., "Large Scale Imputation of Survey Data". Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978. pp. 431-436. Also, Survey Methodology, Statistics Canada, 1978, Vol. 4, No. 2, pp. 203-224.

[3] Cox, B.G. and Folsom, R.E., "An Empirical Investigation of Alternate Item Non-Response Adjustments". Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978, pp. 219-223.

[4] Ernst, L.F., "Weighting to Adjust for Partial Non-Response". Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978. pp. 468-472.

[5] Fellegi, I.P. and Holt, D.A., "Systematic Approach to Automatic Edit and Imputation". Journal of the American Statistical Association, 1976, Vol. 71, pp. 17-35.

[6] Ford, B.L., "Missing Data Procedures: A Comparative Study". Proceedings of the Social Statistics Section, American Statistical Association, 1976, pp. 324-329.

[7] Hill, C.J., "A Report on the Application of a Systematic Method of Automatic Edit and Imputation to the 1976 Canadian Census". Proceedings of the Section on Survey Research Methods, American Statistical Association, 1978, pp. 474-479. Also, Survey Methodology, Statistics Canada, 1978, Vol. 4 pp. 178-202.

[8] Nordbotten, S., "The Efficiency of Automatic Detection and Correction of Errors in Individual Observations as Compared with Other Means of Improving the Quality of Statistics". Bulletin of the International Statistical Institute, Proceedings of the 35th Session, Vol. 16, 1965, pp. 417-441.

[9] Ono, M. and Miller, H.P., "Income Non-Response in the Current Population Survey". Proceedings of the Social Statistics Section, American Statistical Association, 1969, pp. 277-288.

[10] Pritzker, L., Ogus, J. and Hansen, M.H., "Computer Editing Methods - Some Applications and Results". Bulletin of the International Statistical Institute. Proceedings of the 35th Session, Vol. 16, 1965, pp. 442-465.

[11] Sande, G., "Numerical Edit and Imputation". International Association for Statistical Computing, 42nd Session of the International Statistical Institute. December 1979.

[12] Spiers, E.F. and Knott, J.J., "Computer Method to Process Missing Income and Work Experience Information in the Current Population Survey". Proceedings of the Social Statistics Section, American Statistical Association, 1969, pp. 289-293.

[13] Szameitat, K. and Zindler, H.J., "The Reduction of Errors in Statistics by Automatic Corrections". Bulletin of the International Statistical Institute. Proceedings of the 35th Session, Vol. 16, 1965, pp. 395-417.

[14] U.S. Bureau of the Census. 1970 Census of Population and Housing, Procedural History. Chapter 15, Appendix A.

The Editorial Board wish to thank Mr. D.A. Worton, Assistant Chief Statistician, Marketing Services Field, for his continuing support of the publication of this Journal.

Thanks are also due to Miss D.E. Weisenberg, for her patient typing of the Journal, as well as for the execution of numerous other duties associated with its production.

Finally, the Editorial Board wish to thank the following persons who have served as referees during the past year.

Le comité de rédaction désire remercier M. D.A. Worton, Statisticien en chef adjoint, Secteur des services de diffusion, pour son appui constant de la publication de cette revue.

On remercie également Mlle D.E. Weisenberg, qui a dactylographié soigneusement la revue, et qui s'est acquittée de nombreuses autres tâches associées à sa publication.

Le comité de rédaction désire enfin remercier les personnes suivantes, qui ont bien voulu faire la critique des articles présentés au cours de l'année dernière.

D.A. Binder
R.G. Carter
P.D. Ghangurde
A.R. Gower
M. Gratton
G.B. Gray
M.A. Hidiroglou
M.J. Hill
T.M. Jeays
K.P. Krotki
M.C. Lawes
I. Macredie
M.J. March
I.G. Sande
A. Satin
K.P. Srinath
R. Sugavanam
L.A.J. Swain
D.A. Worton

DATE DUE
DATE DE RETOUR

JUN JUIN 12 1980			
AUG AOUT -4 1981			
JUL JUIL -5 1982			
JUN 15 [illegible]			
NOV -6 1990			
[illegible]			

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE

June/juin 1979 Vol. 5 No. 1

A Journal produced by Methodology Staff, Statistics Canada.

Preparé par les méthodologistes de Statistique Canada.

CONTENTS

SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE

December/décembre 1978 Vol. 4 No. 2

A Journal produced by Statistical Services Field, Statistics Canada.

Publié par le secteur des services statistiques, Statistique Canada.

CONTENTS