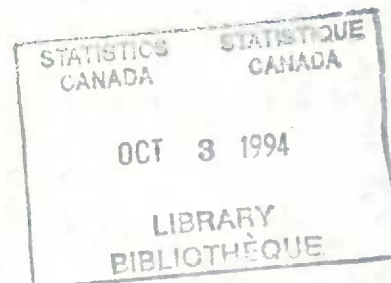# Analytical Studies Branch

Research Paper Series

**LINKING SURVEY AND ADMINISTRATIVE DATA
TO STUDY DETERMINANTS OF HEALTH**

by

Pierre David, Jean-Marie Berthelot et Cam Mustard

No. 58

Health Analysis and Modeling Group
Analytical Studies Branch
Statistics Canada
1993

The analysis presented in this paper is the responsibility of the authors and does not necessarily represent the views or policies of Statistics Canada.

Aussi disponible en français

Linking Survey and Administrative Data to Study Determinants of Health

PIERRE DAVID [1], JEAN-MARIE BERTHELOT [2],
CAM MUSTARD [3], ScD

ABSTRACT

Current health research is finding a very wide range of factors to affect health and health care utilization. Such work is confirming the long-observed relationship between socioeconomic factors and health and expanding understanding of the specific processes underlying the relationship. This paper describes a pilot project that will bring together for the first time in Canada detailed cross-sectional data on health and socioeconomic status with comprehensive longitudinal information on the utilization of health care services for a representative sample of a provincial population. The paper focuses on applications of probabilistic record linkage methods used in combining census and administrative data sources.

KEY WORDS: Probabilistic record linkage; Census and survey data; Socioeconomic status; Health status; Confidentiality.

1. INTRODUCTION

A number of studies have shown a clear relationship between the socioeconomic status of a person and the probability of dying in a given period of time (e.g., Wolfson et al. 1993, Marmot 1986, Wilkins et al. 1991). Other studies have established a link between prevalence of some diseases and socioeconomic characteristics of the neighborhood (Anderson et al. 1993, Dougherty et al. 1990, Gentleman et al. 1991). In addition, cross-sectional Canadian survey data sets have provided information on socioeconomic status and health status, as well as limited data on health care utilization. However, there exists no data base in Canada containing comprehensive longitudinal information about health, health care utilization and socioeconomic status of individuals. In consequence, a pilot project, jointly managed by Statistics Canada and the Manitoba Centre for Health Policy and Evaluation, was launched to examine the feasibility of creating such a data base *from existing data sources.*

The objective of the pilot project is first to evaluate the feasibility of combining information from the 1986 Census of Population, the 1986-1987 Health and Activity Limitations Survey (HALS),

[1] Social Surveys Methods Division, Statistics Canada, Ottawa, K1A 0T6

[2] Social and Economic Studies Division, Statistics Canada, Ottawa, K1A 0T6

[3] Manitoba Centre for Health Policy and Evaluation, Department of Community Health Sciences, Faculty of Medicine

and the longitudinal (1972-1992) Manitoba Health Services Commission (MHSC) health care utilization file. The resulting linked data base will then be used as the basis for significant new research into the determinants of health.

The 1986 Census will provide detailed socioeconomic information such as family composition, dwelling characteristics, occupation, ethnic origin, mother tongue, and income and education related variables. The 1986-1987 HALS is a post-censal survey targetted at Canadians who, for health-related reasons, are limited in the kind or amount of activity they can perform on a day-to-day basis. It will provide information on overall health and activity limitations in addition to employment, education, transportation, housing and leisure activity. Since the health related information in HALS is self declared, it represents the respondent's perceived health status rather than clinical health status. The longitudinal MHSC health care utilization file provides information on hospital visits, diagnoses, surgery, personal care home, date and cause of death, and other related health care utilization data. It has been used for a number of innovative health services research studies (e.g., L.L. Roos et al. 1987, N.P. Roos et al. 1987, Shapiro et al. 1984).

Before initiating the linkage of these data files, a number of procedures were undertaken, following policies in the collaborating agencies. This included consultation with Canada's Privacy Commissioner, the Faculty Committee on the Use of Human Subjects in Research of the University of Manitoba, and Statistics Canada's Confidentiality and Legislation Committee. In addition, the Access and Confidentiality Committee of the Manitoba Health Services Commission was informed of the project.

Following these consultations and the formal policies of Statistics Canada, the Minister responsible for Statistics Canada authorized the linkage project on the terms proposed: it is a pilot project that assesses the feasibility and usefulness of the proposed linkage; names and addresses will not be used to match individual records; the linkage will be performed entirely within the physical premises of Statistics Canada by employees who have sworn the Oath of the Statistics Act; a pilot sample of 20,000 linked records will be used for research and analysis purposes; and access to the linked data will be clearly limited as per the Statistics Act. In addition, all activities with the linked data set are covered by a memorandum of understanding among Statistics Canada, the University of Manitoba and the Manitoba Ministry of Health.

2. METHODOLOGY

The objective of the matching phase of the project is to locate individuals common to the three data sources in order to create, in a subsequent phase of the project, a person-oriented database for a sample of 20,000 linked records. Statistics Canada's Canlink system is used for the matching phase. Canlink is a statistical matching software package that uses the discriminating power of individual-level variables to match records from two separate data files. The comparison of values of conceptually or theoretically identical variables from two records yields a weight that takes into consideration the degree of agreement of the values as well as the probability that an agreement occurs at random. The underlying method builds on the Fellegi-Sunter theory (1969). This section

describes the methodology used for matching a sample of person-based records on the census file (the 2B sample of the 1986 census covering the province of Manitoba) to a subset of the complete file of individual citizens of Manitoba registered with the Manitoba Health Services Commission in June 1986.

2.1 Data Sets

The sample of the 1986-1987 HALS was drawn from the census 2B sample (Dolson et al. 1987). As a result, all records from this data set are already matched to the census data base. Therefore, the only two data files involved in the matching phase of the project are:

1. A subset of variables from the 2B sample of the 1986 population census. This is the long form version of the census questionnaire. It is distributed to approximately twenty percent of all Canadian households. The variables used for the matching phase are the following: residential postal code, month and year of birth, sex, family size, family structure (i.e. single adult or couple, with or without children), family status (grandchild, child, married or common law spouse, parent), mobility between the 1981 and 1986 censuses, and native status. Note that name and street address are not used.

2. The registration file of the MHSC. This file represents all citizens of the province of Manitoba registered with the universal health insurance program as of June 1986. The registration file contains information on registrant year and month of birth, sex, family structure and residential postal code. Because the file is longitudinal, it can be used to describe geographic mobility and family structure changes over time. The registration file has been found to be equivalent to the census as a source of accurate information on population size and structure (Roos et al. 1993), as indicated by the following graph.

Figure 1. 1986 Census 2A Counts Vs Manitoba Registration File Counts By Age Group
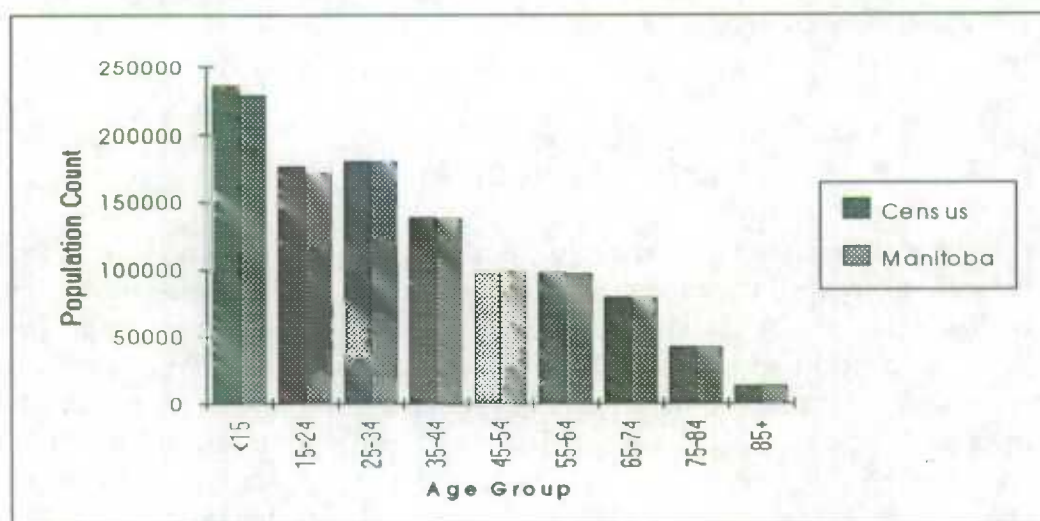
2.2 Pair Forming

The source files for the matching process are derived from the 2B census file containing 261,861 records of individuals living in Manitoba, and from the Manitoba registration file containing 1,047,443 individual records. The number of logically possible pairs of records that can be formed by taking one record from each file is the product of these two quantities, namely over 274 billion. The matching phase consists of identifying the good pairs, that is to say pairs for which records from each file refer to the same individual. Once the good pairings are established, a sample of 20,000 will be drawn to constitute the basis of the linked data base to which we will attach the appropriate analytic variables from each of the source data sets.

Forming and evaluating 274 billion of pairs would be very expensive. In addition, this huge set of pairs would contain at most 261,861 valid pairs, constituting less than 0.0001% of the total. It would be operationally inefficient to form this file and examine all possible pairs. The strategy used to identify good pairs consists instead of dividing the two data sets into identically defined blocks (also called pockets) and forming pairs only from records that belong to the same block.

After examining various possibilities for block definitions, we defined a block in terms of four individual characteristics: sex, month of birth, year of birth and postal code. This means that all the pairs considered for matching are formed from individuals who agree exactly on these four variables. This yielded a great number of small blocks, each containing between 1 and 22 records. This two step strategy of forming blocks first and then examining prospective pairs is more efficient in terms of matching then simply evaluating all possible pairs.

2.3 Pair Weighting

For each pair, selected variables are compared one at a time and a weight proportional to the degree of agreement is given. Based on calculated and a priori probabilities, perfect agreements receive high weights, disagreements receive low weights, and partial agreements, when used, receive intermediate weights. Afterwards, the sum of these weights, called comparison weights, gives the total weight of a pair. This total weight reflects the likelihood that the pairing is good. In other words, the total weight is proportional to the probability that the records forming the pair belong to the same individual.

The comparison weights are calculated from the odds ratio of conditional probabilities of possible outcomes (agreement, partial agreement, disagreement) among true matches and true non-matches (Statistics Canada 1989, David 1992). For comparison variable i and outcome j, the odds ratio is defined as:

$$(1) \quad R_{ij} = \frac{P(Outcome\ j\ |\ true\ match)}{P(Outcome\ j\ |\ true\ non-match)}$$

4

Since it is more convenient to use additive functions and integers, the comparison weight for variable i and outcome j is defined as:

(2) $W_{ij} = INT(\ 10 \times LOG_2\ (R_{ij})\ )$

The odds ratio for a specific pair is obtained by multiplying the odds ratios R_{ij} over all comparison variables given the observed outcomes for that pair. In consequence, the total weight of a specific pair is the sum of all comparison weights given the observed outcome for that pair.

Agreements which are more likely among matches than among non-matches receive positive weights since the odds ratio R_{ij} is greater than 1 in this case. By construction, variables which have a high number of response levels have a higher discriminating power and generate larger agreement weights. This can be seen in the table 1, where the denominator of R_{ij} is estimated by the probability of outcome j among pairs formed at random. The numerator of the odds ratio is estimated iteratively using samples of pairings which are deemed to be true matches. Numerators shown in table 1 are used to illustrate weight calculations only and are not actual estimates of the corresponding probabilities.

Table 1. Agreement and Disagreement Weights for Child Sex and Child Month of Birth

	Sex	Month of Birth
$R_{iA} = \dfrac{P(Agreement \mid true\ match)}{P(Agreement \mid true\ non-match)}$	$\dfrac{0.984}{1/2} = 1.97$	$\dfrac{0.980}{1/12} = 11.76$
Agreement Weight = INT $(10 \times LOG_2\ (R_{iA}))$	9	35
$R_{iD} = \dfrac{P(Disagreement \mid true\ match)}{P(Disagreement \mid true\ non-match)}$	$\dfrac{0.016}{1/2} = 0.03$	$\dfrac{0.020}{11/12} = 0.02$
Disagreement Weight = INT $(10 \times LOG_2\ (R_{iD}))$	-49	-55

A few variables are used in table 2 to illustrate how the weights are applied. More variables are actually used in weighting the pairs. In this example, since both individuals are declared to be married, a perfect agreement weight is given for variable marital status. A lower weight is given for partial agreement on the variable family size. The spouse year of birth agrees and a higher weight is given, reflecting that this variable is more discriminating than the marital status. A negative weight is given for disagreement on the spouse's month of birth. Finally, the sum of the comparison weights gives this pair a total weight of 25.

Table 2. Weighting Example

	Marital Status	Family Size	Spouse Year of Birth	Spouse Month of Birth	Total
Census	married	4	1956	10	
Manitoba	married	3	1956	4	
Weight	9	2	45	-31	25

After examining the content of both data sets, the following variables were selected for pair weighting: marital and native status of the individual; month and year of birth of the spouse; sex, month and year of birth of the youngest child (if any); and finally, size, structure and geographic mobility of the family.

Once all pairs are weighted, they are classified into three groups according to their total weight: a reject group, a possible group and a definite group. Thresholds are used to define the classification groups and are determined by examining a representative sample of weighted pairs. They should divide the pairs into three relatively homogeneous groups. If the weighting is appropriate, then the pairs will be arranged in ascending order according to the likelihood that they are good pairs. Figure 2 illustrates a fictitious but typical case.

Figure 2. Pair Distribution According to Total Weight
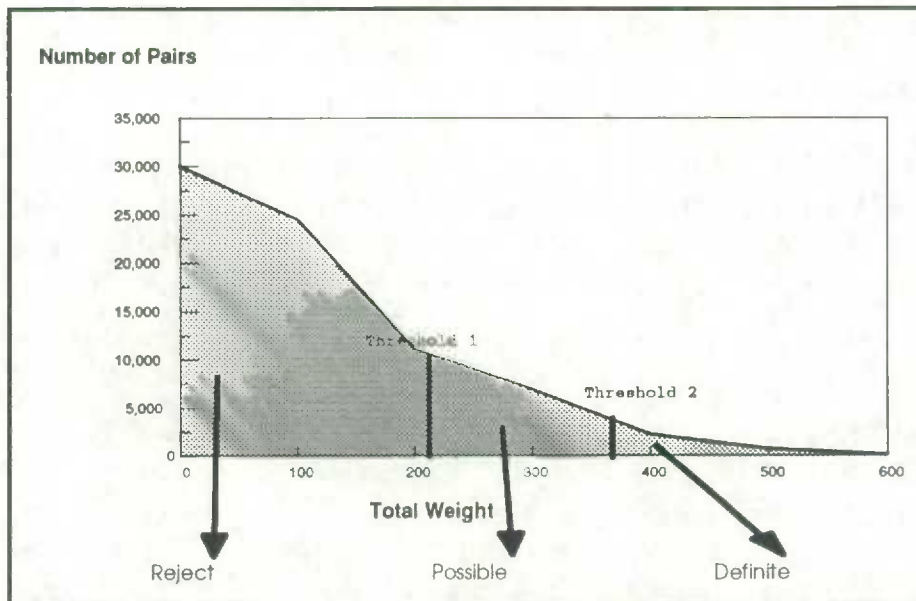
A high proportion of pairs have a total weight that falls below threshold 1. These pairs are made up of individuals who agree on block variables, but disagree on most or all comparison variables.

The majority of them are not good and can be rejected confidently. The principal objective of this weighting is to reject pairs that are manifestly bad. Threshold 1 plays the most important role in this: if placed too low, then some bad pairs are likely to be accepted as possible pairs, if placed too high, then some possibly good pairs are likely to be rejected.

For a small number of pairs, most variables agree, indicating a high probability that individuals are the same. The majority of these pairs are clearly good and their weight is above threshold 2.

The validity of pairs for which the total weight lies between thresholds 1 and 2 is uncertain. Some of the pairs are not good: they are made of individuals who have very similar characteristics but are not the same. The other pairs are good but they contain errors that yield a lower weight than they should get. In general, data errors, updating errors and conceptual errors introduce noise that usually makes two records of a unique individual look different. Let us briefly describe these three types of errors.

1. Inaccurate data reported by the respondent and capture errors are examples of data errors: the individual reports a year of birth of 1954 instead of 1953, or the month of birth is keyed in as 12 instead of 2. Although capture techniques can be very sophisticated and efficient, this type of error is hard to eliminate completely and can be considerably misleading. For example, if the year of birth is part of the block variables, then the records won't even be compared, unless the same error appears on both files, which is highly unlikely.

2. Updating errors occur when data are collected or updated at different times. For example, the census collects data on a specific date (June 3, 1986), while the information residing on the Manitoba registration file is usually updated every six months. Different reference dates inevitably cause data discrepancies. For example, someone can be declared single on the census and be married on the Manitoba file if the marriage occurred between the census date and the Manitoba update.

3. The third type of error deals with the conceptual frameworks inherent in the data bases to be matched. For example, the census and the Manitoba registration file use different definitions of a family. Even though census data were recoded to match as closely as possible the Manitoba definition, some discrepancies may remain, reducing the probability of establishing true matches.

2.4 Frequency Weighting

The objective of the frequency weighting is to improve pair ordering by using weights that are proportional to the rarity of the value on which records agree. This type of weighting is computationally more expensive than the first one since it associates a specific weight to every agreement value. Hence it is used only after the numerous pairs that were obviously bad have been rejected.

Table 3. Example of Frequency Weighting

	Marital Status	Family Size	Spouse Year of Birth	Spouse Month of Birth	Total
Census	married	2	1944	10	
Manitoba	married	2	1944	10	
Weight	5	20	50	35	110

	Marital Status	Family Size	Spouse Year of Birth	Spouse Month of Birth	Total
Census	married	10	1899	12	
Manitoba	married	10	1899	12	
Weight	5	100	170	35	310

In table 3, agreement on a family size of ten gets more weight than agreement on a family size of two. Also, agreement on a rare year of birth generates more weight than agreement on a more common one. Variables like the month of birth, for which all values are deemed to be reasonably equiprobable, receive a fixed weight.

The frequency weighting generates a new pair distribution as illustrated in figure 3. New thresholds are then determined from a sample of weighted pairs to obtain the final pair classification.

Figure 3. Pair Distribution After Frequency Weighting

8
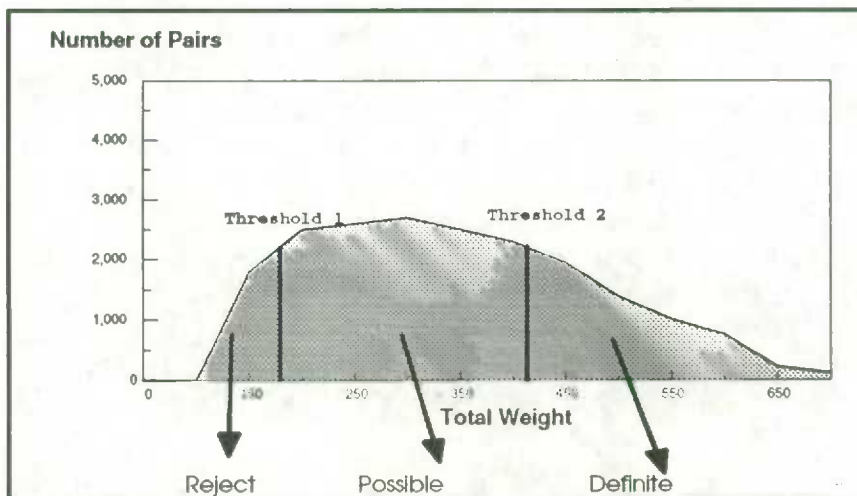
3. RESULTS

Overall, 70.4% of individuals from the census file were matched on a one-to-one basis to individuals from the Manitoba file. Figure 4 shows little difference between men and women. Note that approximately 6% of individuals belong to pairs classified as "possible". The records which constitute such pairs may refer to the same individual, though the pairing has a lower weight due to data errors, or could relate to similar but different persons. The relatively small 6% figure indicates that we preferred to limit the number of possible pairs by using a relatively high threshold 1. Usually, this strategy rejects a few good pairs, but in return, allows few bad pairs to be kept. Hence the quality of the selected pairs should be relatively good.

Figure 4. Match Rate of the Census File
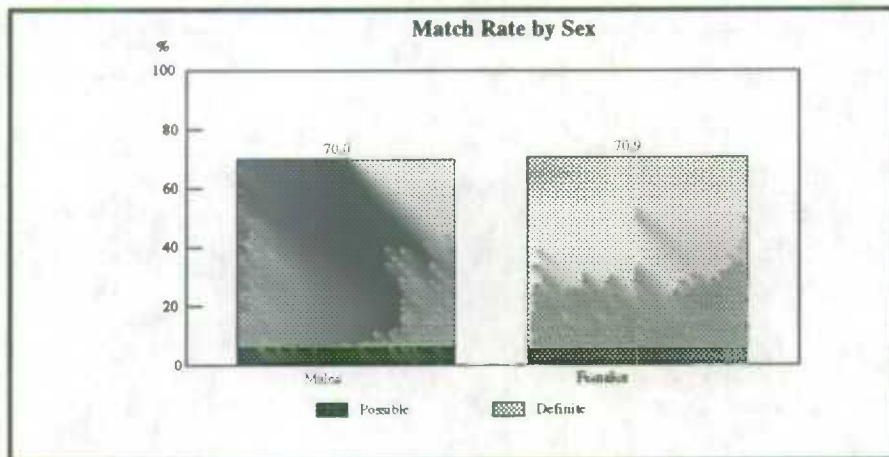
3.1 Mobility

The major factors affecting the match rate are related to the geographic mobility of individuals. For instance, the following groups were harder to match: young adults (20 to 25 years of age), people who changed dwelling between the 1981 and the 1986 censuses, and people who are either separated or divorced. Within these groups, frequent address changes as well as family structure changes make concordance between data sources more difficult than for less mobile groups. In fact, since census data are dated June 3, 1986, and since most Manitoba variables are dated December 31, 1986, a lag in the data is more probable among mobile individuals. Figure 5 illustrates match rates according to some of these variables.

Figure 5. Census File Match Rates According to Various Variables
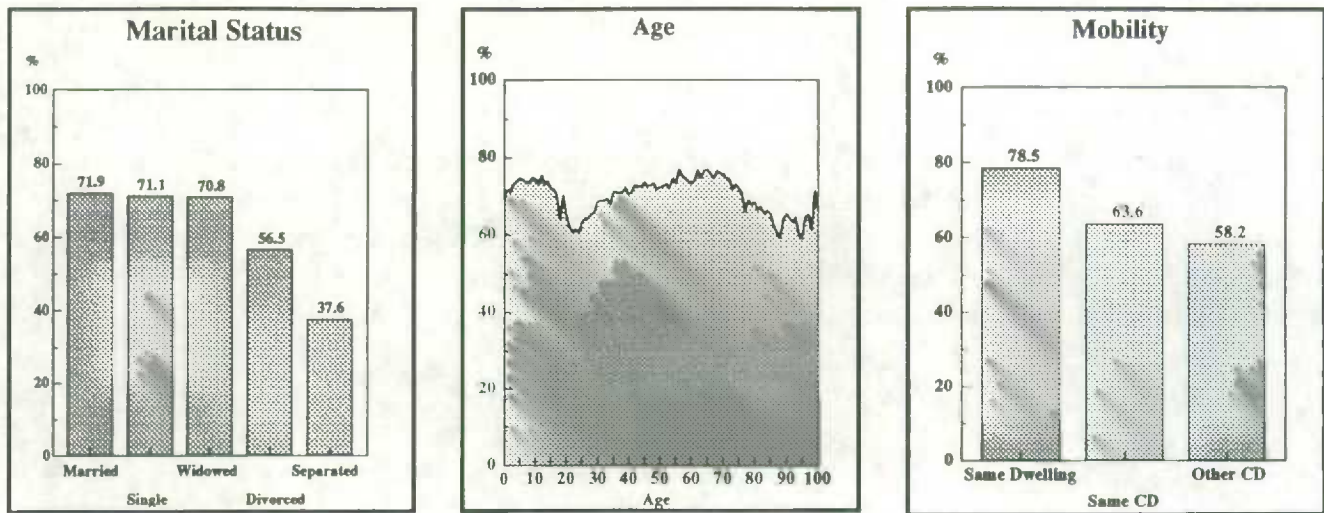
CD: Census division, a geographic area used by the census of population. The province of Manitoba contains 23 census divisions.

A low match rate is observed among separated people. It can be explained by the mobility inherent in the separation phenomenon, as well as by the lag between data sources.

The effect of age on the match rate is not surprising. Children under 15 and adults between 30 and 60 years of age get better rates given their more stable situation. Due to institutionalization and a limited number of cases, more variability is observed among people over 85.

We could have expected an even better match rate for people who did not move between the 1981 and the 1986 censuses (same dwelling). The 78.5% rate observed with this group may suggest that the maximum match rate, given the data errors in both files, is around 80% when using the current methodology.

3.2 Postal Code

On the census file, the postal code is dated June 3, 1986. Six percent of the records had a missing postal code. In these cases, Statistics Canada's Geography Division derived postal codes using a high quality procedure based on the relationship between census geography and postal codes. The use of this derived postal code yielded good results, contributing 4% of total matches.

On the Manitoba file, postal codes are dated December 31 of each year. To match the files, we used the 1986 postal code as the basic postal code. We also used three alternative postal codes from the Manitoba file: the 1985 vintage, the 1987 vintage, and a second 1986 postal code for individuals who had an alternative address in 1986. The use of alternate postal codes provided 7%

of total matches. In conclusion, these two methods were useful, generating matches that would have been missed otherwise.

3.3 Family Reconciliation

After 70.4% of the census records were confidently paired through the use of Canlink, we examined census and Manitoba families for which all members but one had been matched. When the unmatched members of the corresponding families were alike (same sex and same age to within 5 years), we paired them into definite matches. This procedure added almost 2% more matches, increasing the global match rate to 72.1%.

4. PASS TWO

Even though the 72.1% match rate is fairly good, the relatively different characteristics of unmatched individuals argued in favour of a second match wave. The analysis of pass two has not yet been completed.

4.1 Preliminary Work

For pass two, all individuals not matched in pass one were included, as well as some individuals for which the match was not entirely satisfactory. This is the case for the following groups:

1. People living alone for which the match was classified as possible (3,390 census records). The validity of these matches is difficult to establish since very few variables are effectively compared.

2. Incomplete families (62,888 census records). All family members were included in pass two if at least one person in the family was not matched in pass one.

3. Some complete families (13,076 census records). In census families for which members were matched to members of more than one Manitoba family, everybody from both data files were included in pass two.

In pass one, an individual had to agree exactly on each of the four block variables (sex, month of birth, year of birth and postal code) to be compared to a counterpart and possibly matched. A single error on the month of birth for example would have prevented a good pair from being formed.

In pass two, the block definition was enlarged to allow for more pairs to be formed. The exact month and year of birth were replaced by the age of the person. This allowed a record to be compared to more potential candidates. Moreover, the area covered by the geographic variable

was enlarged in urban areas as we substituted the census enumeration area for the postal code (i.e. about two to three times as large an area).

With regard to comparison variables, we used the same variables as in pass one, with the following exceptions: native status was discarded due to definitional problems between data sources; the month and year of birth were used as comparison variables; in urban areas, the postal code was used as a comparison variable; finally, the family structure was redefined to make definitions of grandchildren and common law unions more comparable.

4.2 Results

Overall, 45% of census individuals included in pass two were matched to a single individual from the Manitoba file. Considering that the best matches had been formed in pass one and that they were excluded from pass two, this rate appears satisfactory.

Among pass two matches, a high proportion of possible pairs occurred in urban areas. As expected, they occurred mainly among young single adults for whom very little information could be compared beside the block variables (i.e. family size is always one, marital status is almost always single, there is no spouse or child information to compare, etc.).

Figure 6 shows the match rate of the census file for pass 1 and 2, as well as the expected final results of both passes.

Figure 6. Census File Match Rates

		PASS 1	PASS 2	GLOBAL
DEF		66.4	10.8	71.4
POS		5.7	7.2	9.0
REJ		27.9	22.1	19.6
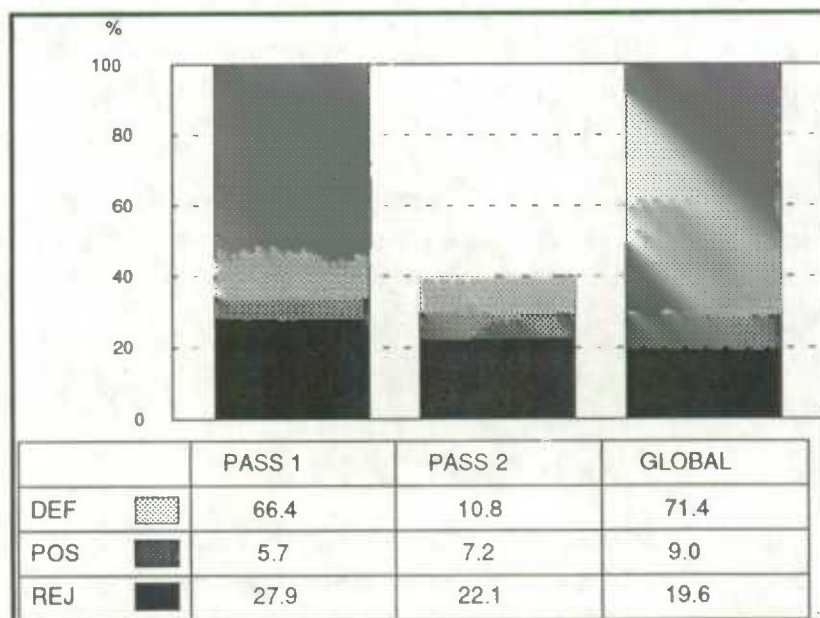
5. CONCLUSION

In conclusion, the methodology presented in this article allows approximately 80% of the census file (71.4% + 9.0%, see figure 6) to be statistically matched to the Manitoba registration file, based essentially on age, sex, postal code, family size and family structure. A few refinements are left in terms of matching which could raise this rate by one or two percentage points. For example, we could easily establish matches in families for which only one member remains not matched (as done after pass 1). This work, as well as the analysis of pass two matches, is next on the agenda and will complete the matching phase of the project.

The 80% rate is satisfactory in comparison with typical survey response rates. For example, the Nova Scotia Nutrition Survey achieved response rates of 79.7% among located respondents and 60.0% among total sample drawn (Nova Scotia Heart Health Program). The Manitoba Heart Health Survey achieved response rates of 77.1% among located respondents and 60.8% among total sample drawn (Young et al.).

Clearly, when considering the various types of discrepancies that can afflict statistical matching, a 100% rate becomes unrealistic. Data errors, lags in data collection or updating, and conceptual differences in the data bases to be matched inevitably limit the success rate of any statistical matching. Here, unmatched individuals present relatively different characteristics than matched individuals. However, very detailed socio-demographic information about the non-matched population is available from the census file. This information will enable us to select a sample for the next analysis phase which will be representative of the whole population.

Future activities include a quality evaluation of the matches obtained. The planned method consists of selecting a sample of one thousand or two thousand matches, and then hand comparing names and addresses from more detailed data not generally available (e.g. the hard copy of the census form). This information would not be used to validate specific matches but only to estimate true match rates at aggregated levels. These rates will then be used to accept or reject entire groups of matched records.

Afterwards, a sample of 20,000 individuals, representative of the population to be studied, will be selected. Health and socioeconomic variables will be added to the matched records and data will be organized into a unique data base that will support analyses of the relationships among socioeconomic status, health and health care utilization.

REFERENCES

ANDERSON, G., GRUMBACH, K., LUTT, H., ROOS, L.L., and MUSTARD, C. (1993). Use of coronary artery bypass surgery in the United States and Canada: influence of age and income. *Journal of the American Medical Association*, 269, 1661-1666.

DAVID, P. (1992). Methods for calculating probabilities and weights. Appendix 1 of internal report dated November 17, 1992, Statistics Canada.

DOLSON, D., McCLEAN, K., MORIN, J.-P., and THÉBERGE, A. (1987). Plan d'échantillonnage pour l'enquête sur la santé et les limitations d'activités. *Techniques d'enquête*, 13(1), 101-117.

DOUGHERTY, G., PLESS, I.B., and WILKINS, R. (1990). Social class and the occurrence of traffic injuries and death in urban children. *Canadian Journal Of Public Health*, 81, 204-209.

FELLEGI, I.P. and SUNTER, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64, 1183-1210.

GENTLEMAN, J.F., WILKINS, R., NAIR, C., and BEAULIEU, S. (1991). An analysis of frequencies of surgical procedures in Canada. *Health Reports*, 3(4), 291-309.

MARMOT, M.G. (1986). Social inequalities in mortality: the social environment. In *Class and Health, Research and Longitudinal Data*, (Ed. R.G. Wilkinson). London: Tavistock Publications.

Nova Scotia Heart Health Program, Nova Scotia Department of Health, Health and Welfare Canada. Report of the Nova Scotia Nutrition Survey.

ROOS, L.L., MUSTARD, C.A., NICOL, J.P., McLERRAN, D.F., MALENKA, D.J., YOUNG, T.K., and COHEN, M.M. (1993). Registries and administrative data: organization and accuracy. *Medical Care*, 31(3), 201-212.

ROOS, L.L., NICOL, J.P., and CAGEORGE, S.M. (1987). Using administrative data for longitudinal research: comparisons with primary data collection. *Journal of Chronical Diseases*, 40(1), 41-49.

ROOS, N.P., MONTGOMERY, P., and ROOS, L.L. (1987). Health care utilization in the years prior to death. *The Milbank Quarterly*, 65(2), 231-254.

SHAPIRO, E. and ROOS, L.L. (1984). Using health care: rural/urban differences among the Manitoba elderly. *The Gerontologist*, 24(3), 270-274.

Statistics Canada, System Development Division (1989). Generalized iterative record linkage system weights.

WILKINS, R., ADAMS, O., and BRANCKER, A. (1991). Changes in mortality by income in urban Canada from 1971 to 1986. *Health Reports*, 1(2), 137-174.

WOLFSON, M.C., ROWE, G., GENTLEMAN, J.F., and TOMIAK, M. (1993). Career earnings and death: a longitudinal analysis of older Canadian men. *Journal of Gerontology: Social Sciences*.

YOUNG, T.K., GELSKEY, D.E., MACDONALD, S.M., HOOK, E., and HAMILTON, S. The Manitoba heart health survey: technical report.

ANALYTICAL STUDIES BRANCH
RESEARCH PAPER SERIES

No.

17. *Wages and Jobs in the 1980s: Changing Youth Wages and the Declining Middle,* **J. Myles, G. Picot and T. Wannell**

18. *A Profile of Farmers with Computers,* **Ray D. Bollman**

19. *Mortality Risk Distributions: A Life Table Analysis,* **Geoff Rowe**

20. *Industrial Classification in the Canadian Census of Manufactures: Automated Verification Using Product Data,* **John S. Crysdale**

21. *Consumption, Income and Retirement,* **A.L. Robb and J.B. Burbridge**

22. *Job Turnover in Canada's Manufacturing Sector,* **John R. Baldwin and Paul K. Gorecki**

23. *Series on The Dynamics of the Competitive Process,* **John R. Baldwin and Paul K. Gorecki**

    A. *Firm Entry and Exit Within the Canadian Manufacturing Sector.*
    B. *Intra-Industry Mobility in the Canadian Manufacturing Sector.*
    C. *Measuring Entry and Exit in Canadian Manufacturing: Methodology.*
    D. *The Contribution of the Competitive Process to Productivity Growth: The Role of Firm and Plant Turnover.*
    E. *Mergers and the Competitive Process.*
    F. *(in preparation)*
    G. *Concentration Statistics as Predictors of the Intensity of Competition.*
    H. *The Relationship Between Mobility and Concentration for the Canadian Manufacturing Sector.*

24. *Mainframe SAS Enhancements in Support of Exploratory Data Analysis,* **Richard Johnson and Jane F. Gentleman**

25. *Dimensions of Labour Market Change in Canada: Intersectoral Shifts, Job and Worker Turnover,* **John R. Baldwin and Paul K. Gorecki**

26. *The Persistent Gap: Exploring the Earnings Differential Between Recent Male and Female Postsecondary Graduates,* **Ted Wannell**

27. *Estimating Agricultural Soil Erosion Losses From Census of Agriculture Crop Coverage Data,* **Douglas F. Trant**

28. *Good Jobs/Bad Jobs and the Declining Middle: 1967-1986,* **Garnett Picot, John Myles, Ted Wannell**

29. *Longitudinal Career Data for Selected Cohorts of Men and Women in the Public Service, 1978-1987,* **Garnett Picot and Ted Wannell**

For further information, contact the Chairperson, Publications Review Committee, Analytical Studies Branch, R.H. Coats Bldg., 24th Floor, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, K1A 0T6, (613) 951-8213.