C.8

12999

COVERAGE ERROR IN THE CANADIAN LABOUR FORCE SURVEY

STATISTICS CANADA STATISTIQUE CANADA

MAY 3 1993

LIBRARY BIBLIOTHÈQUE

Colleen Clark, Brian Kennedy, Marek Wysocki
Statistics Canada

ABSTRACT

The Canadian Labour Force Survey (LFS) uses a multi-stage design consisting of an area frame at the earlier stages and a list frame of dwellings at the final stage. Coverage errors occur when the target population is not adequately represented at any stage of sample selection. The LFS has historically undercovered its target population by four to six percent relative to the Canadian Census. Trends such as the tendency to miss young males will be reviewed. The factors that contribute to coverage error will be discussed via information from ongoing monitoring programs, special studies, and Census results. Ongoing efforts to deal with coverage error are presented including the application of independent post-censal estimates to reduce bias and the use of field procedures designed to minimize coverage error.

KEYWORDS

Survey Coverage Error, Survey Undercoverage, Post-Censal Estimates of Population, Nonsampling Error, Data Quality

1. INTRODUCTION

The LFS is a monthly survey from which standard measures of Canada's labour market conditions are produced including estimates of employment, unemployment, and labour force participation and the corresponding rates. The survey design is multi-stage with an area frame at the earlier stages of selection and a sample of dwellings from lists created by field staff at the final stage. Interviewers determine members of the target population within selected dwellings at the point of first contact.

Coverage error occurs when the target population is not adequately represented at any stage of sample selection. The LFS target population is all persons 15 years of age and over who reside in the 10 provinces of Canada who are either Canadian citizens or landed immigrants. The following are excluded: inmates of institutions, full-time members of the Canadian Armed Forces, and those living on Indian reserves. The exclusions represent about 2% of the population.

A measure of coverage error, referred to as the "slippage rate," is produced for each monthly occurrence of the survey. It is based on the difference between estimates of the target population from the survey and independent post-censal estimates of the same population derived from administrative data and Census counts unadjusted for Census coverage error:

$$\text{Slippage} = 100 * (P - D)/P$$

D = design-based survey estimate
P = post-censal estimate

Slippage rates are produced monthly for Canada, by province and by age-sex group. These rates are monitored by the LFS Quality Committee which meets every month prior to the release of the survey estimates to review all of the survey's quality measures including nonresponse, slippage, interviewer turnover, vacancy (dwellings declared out-of-scope), and data capture error.

The LFS has historically been subject to 4-6% undercoverage. In February 1993 undercoverage was estimated to be 7.1% representing 1.5 million people in a population of 21.3 million. This rate is one of the highest points in the series since 1985. Survey undercoverage varies by age and sex and by geographic area. In order to control bias in the design-based estimates resulting from this undercoverage they are adjusted to post-censal estimates of the target population (the same estimates used in calculating slippage) by age group, sex, and subprovincial region using a regression estimation outlined in LeMaître and Dufour (1987).

Undercoverage is caused when either households or individuals within households that should have been included in the survey are missed. Households can be missed when selecting from the area frame, and in creating the list frame of dwellings, while individuals can be missed once an interviewer has contacted a household.

Coverage error, relative to other measures of survey quality, is difficult to understand and control since it is influenced by a number of factors. Many of these factors are minor in their influence. Further, good measures of coverage error are rare and costly. The slippage measure used by the LFS, for example, is highly influenced by the survey design, field operations, and, to a lesser extent, by the benchmarks taken to represent the "true" number of individuals in the population. While it is possible to identify factors contributing to slippage, the relative impact of these factors and their interactions are not fully understood. Thus, one is left at somewhat of a loss to explain why slippage went up or down from one month to the next.

This paper gives levels of coverage error and describes possible causes of that error for the Canadian Labour Force Survey (LFS). The limitations of the measure used to monitor coverage error are also discussed. A review of the survey's design is in the next section. The field procedures relevant to coverage error are then described. Historical trends of coverage error follow. An overview of the factors which have a potential for influencing coverage error and the slippage measure are described. The paper then focuses on three of these factors. The behaviour of the slippage measure as explained by selection from the area frame is described. The opportunity to miss individuals once an interviewer enters a household is then discussed. Last, the methodology of the post-censal estimates and an evaluation of their role as benchmarks for the calculation of slippage rates are given. Conclusions, and recommendations for both further study and research to reduce coverage error and improve it's measurement complete the paper.

2. SURVEY DESIGN

The Labour Force Survey is a monthly survey of about 63,000 households. A complex multi-stage design is used based on 6 rotation groups. Households are in the survey for 6 months. In any one month 1/6 of the households are in their first month, 1/6 in their second month, and so on. Households are visited in the first month and telephoned in the subsequent months.

Each province is delineated into sub-provincial regions for which labour market data is required. Within each sub-provincial region, three non-overlapping frames are used - namely the Self Representing (SR) frame, the Non Self Representing frame (NSR), and the Special Areas frame.

The SR frame consists of all large urban centres in the country. In order to qualify a city must have a population sufficient to result in a sample size equal to at least one interviewer assignment or have unique labour force characteristics. In each SR area an optimal stratification is carried out. Within each stratum

a set of well defined areal units or clusters (roughly equivalent to city blocks) are delineated. The clusters are then randomly assigned to one of 6 (or multiple of 6) random groups following the method of Rao, Hartley, Cochran (1962). One cluster is selected from each random group with probability proportional to the 1981 Census count of private occupied dwellings (PPS). All dwellings in the selected cluster are listed via field work. Systematic samples of dwellings are drawn over time until the cluster is exhausted and another cluster is selected to replace it. Clusters rotate over time following a procedure which ensures the PPS sampling.

The NSR design is currently based on the 1981 Census. The frame consists of small urban centres and rural areas not part of the special area frame. An optimal stratification is carried out within each sub-provincial area. When sufficient urban population existed within the sub-provincial area, the stratification was carried out within the rural and urban parts of the area separately. Primary Sampling Units (PSU) consisting of urban centres or geographically contiguous groupings of Census enumeration areas were then formed optimally within each strata. The PSU's are selected using probabilities of selection based on 1981 Census counts (PPS). Either two or three PSU's are selected from each strata. While delineation of the second stage areal units differs in rural and urban areas, sampling is PPS. Rotation of the areal units is similar to that of the SR areas except cost considerations necessitate a minimum retention period for PSU's.

Special frames are constructed for remote regions of the country and for institutions. Multistage PPS sampling is employed. The special area frame covers less than 2% of the population.

More information on the sample design and the rotation pattern can be found in Singh et al. (1990).

3. FIELD PROCEDURES

Listing Dwellings

Once a cluster has been selected for the survey an interviewer lists all dwellings in the cluster using a map sketched by Head Office staff and the following rules: A dwelling must be structurally separate from the living quarters of other dwellings. It must have a private entrance from either the outside or a common hall way. It must not be necessary to pass through the living quarters of another dwelling. Clusters are listed about six months prior to when they are scheduled to rotate into the survey.

A cluster normally remains in the sample for several years. Therefore, it is necessary to maintain the list of dwellings. A listing check is conducted the first month that a cluster is in the sample and every 6 months thereafter. A listing check is also done when the interviewer changes or at the request of senior staff. Further, if the cluster is one in which growth since 1981 is large enough to necessitate sub-sampling (termed a "growth" cluster) a check is done every month. Senior interviewers may also conduct a listing check as part of regular monitoring of interviewers.

Classifying Dwellings

At the first interview it must be determined if each selected dwelling is in-scope for the LFS. A dwelling is out-of-scope if it is determined to be vacant, under construction, non-existent (demolished, converted to business, moved, abandoned, or, listed in error), or occupied by a person or persons not in the LFS target population.

Listing Household Members

A household is any person or group of persons living in a dwelling. Interviewers determine, at the first interview, everyone in a household who is eligible for the survey. This is done by constructing a list of

household members. Household members are individuals who are in the LFS target population and who regard the dwelling to be their usual place of residence or are staying in the dwelling and have no usual place of residence elsewhere. Usual place of residence is determined by the following questions:

WHAT ARE THE NAMES OF ALL PERSONS NOW LIVING OR STAYING HERE WHO HAVE NO USUAL PLACE OF RESIDENCE ELSEWHERE?

ARE THERE ANY PERSONS AWAY FROM THIS HOUSEHOLD ATTENDING SCHOOL, VISITING, TRAVELLING OR IN HOSPITAL WHO USUALLY LIVE HERE?

DOES ANYONE ELSE LIVE AT THIS DWELLING SUCH AS OTHER RELATIVES, ROOMERS, BOARDERS OR EMPLOYEES?

The information given in the first interview is confirmed in subsequent months by asking:

DO THE FOLLOWING PERSONS STILL LIVE OR STAY AT THIS DWELLING?

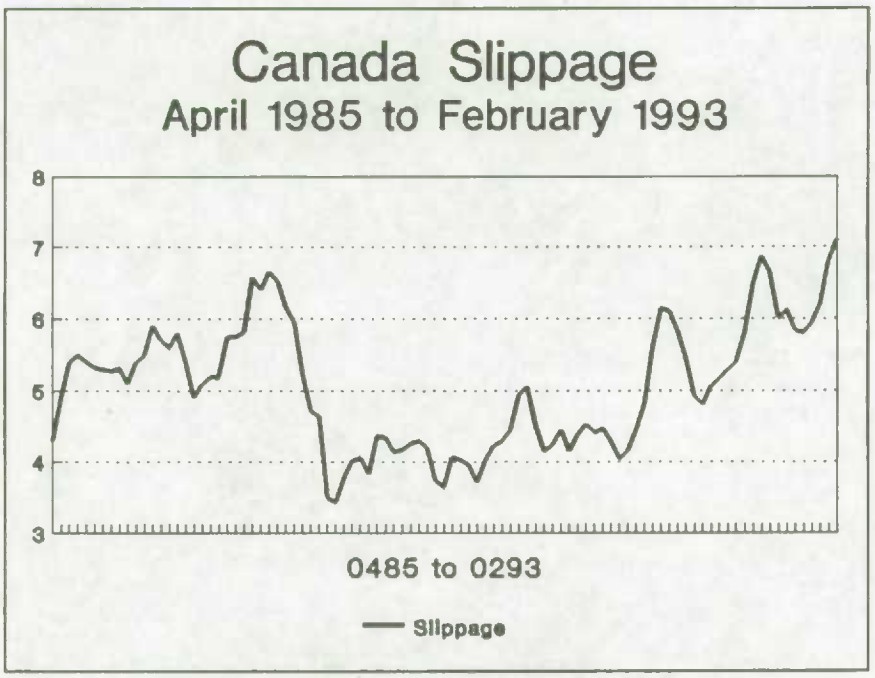
It is usually straight forward to determine usual place of residence but some cases are complex. The LFS Interviewers Manual states:

Should the respondent be unclear about whom to include, follow these rules:

- **Single students are included in the parental dwelling if they have spent 30 or more days there in the past 12 months. Otherwise they are included in the dwelling they are using in survey week.**
- **Married person maintaining or staying in a dwelling other than the spouse's for reasons related to work or school are included in the spouse's dwelling.**
- **Unmarried sons and daughters (non-students) living and working away from the parental dwelling are included in the dwelling used while working.**

4. SLIPPAGE TRENDS

The following graph gives an 8-year series of the slippage rate for Canada from April 1985 (the second month in which a new sample based on the 1981 Census was fully implemented) to February 1993. The estimated standard error of the national slippage rate has fluctuated between about 0.75 and 0.95 (of a percent) since 1989.



Slippage began at 4.3% in April 1985 and quickly moved up to just under 5.5%. It remained around that level for just under two years and then climbed to 6.5% in June 1987. There was a large decrease down to just under 4.5% over the next months where it remained for the following three years to late 1990. Since early 1991, the trend in slippage has been upward.

The large reduction in slippage late in 1987 and into early 1988 was the result of several changes to the Interviewers Manual in April 1987 (Switzer, 1987). The flavour of the revisions were to "include a dwelling when in doubt." This meant, for example, that all dwellings suspected by the interviewer to be seasonal were to be listed thus moving the decision on its inclusion or exclusion from the point of listing to the point at which a contact is made or attempted.

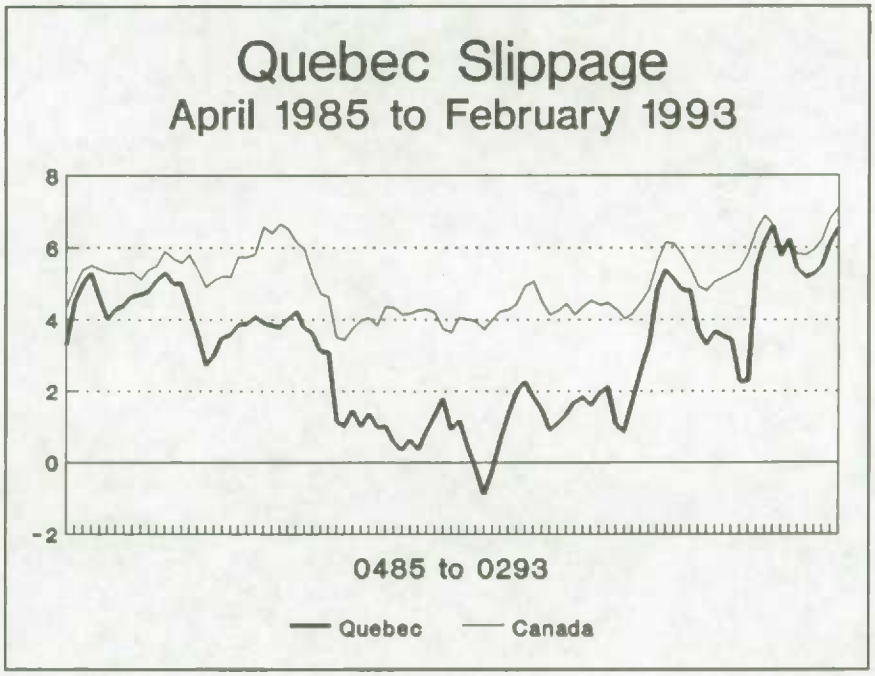
The increasing trend in slippage in recent years reflects the out of date state of the design which is based on counts from the 1981 Census. One motivation for redesigning the survey every ten years is to update these counts with more recent Census counts.

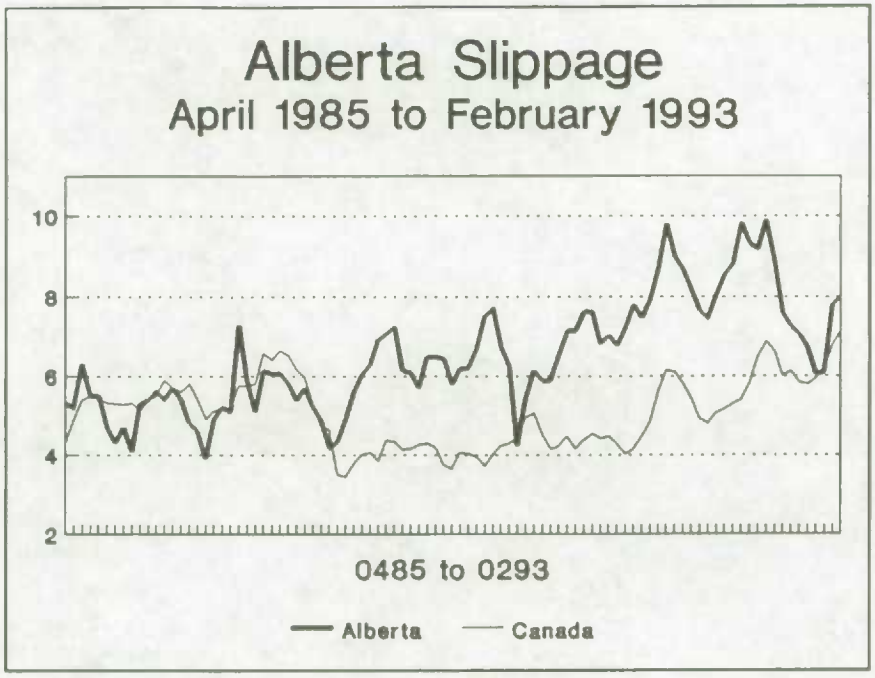
4.1 PROVINCIAL TRENDS

The above graph indicates that the problem of undercoverage, as measured by the slippage rate, has been steadily increasing at the national level since 1991. This nation-wide trend has by no means been echoed in all provinces. The province of Saskatchewan, for example, has undergone a marginal decline in its slippage rate since early 1992. Since 1985, the undercoverage level for Newfoundland has shown an overall decline. Prince Edward Island (P.E.I.) and New Brunswick (N.B.) have also exhibited slippage series that bear little resemblance to the national one. However, two provinces that have followed the national trend over the 1989 to 1993 period are Quebec and Ontario. In fact, national slippage rates are highly influenced, even driven, by those of Quebec and Ontario, which are Canada's two most populous provinces.

While the slippage rates in Quebec from April 1985 to February 1993 and in Ontario from about mid-1990 onwards were, in general, below the Canadian level, most other provinces, in general, have experienced higher undercoverage rates. An example of this is Alberta, a western province of Canada, which over the last five years has experienced higher slippage by as much as 3 or 4 percentage points.

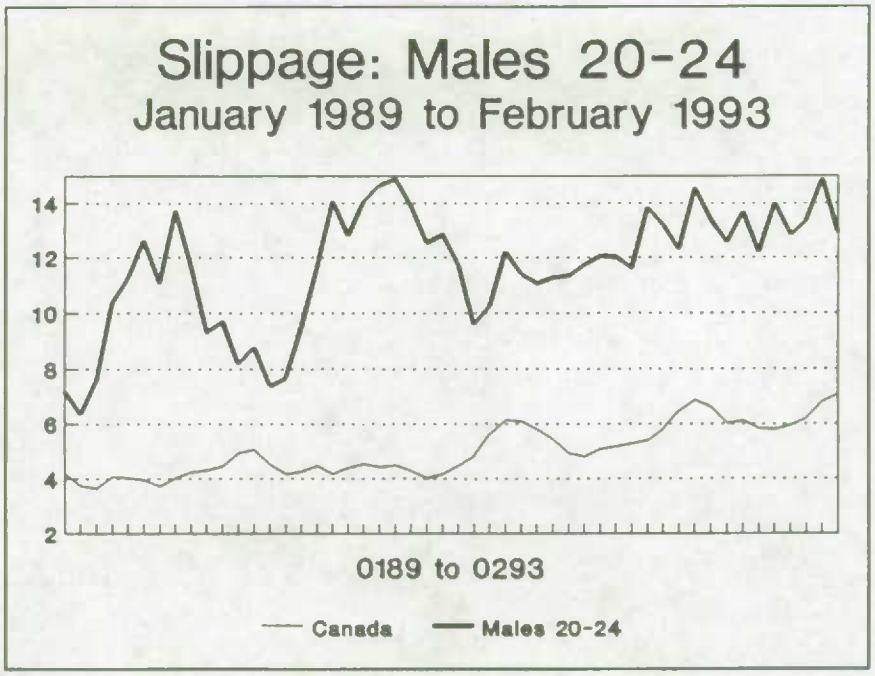
Undercoverage, for Quebec and Alberta, is shown in the following graphs.

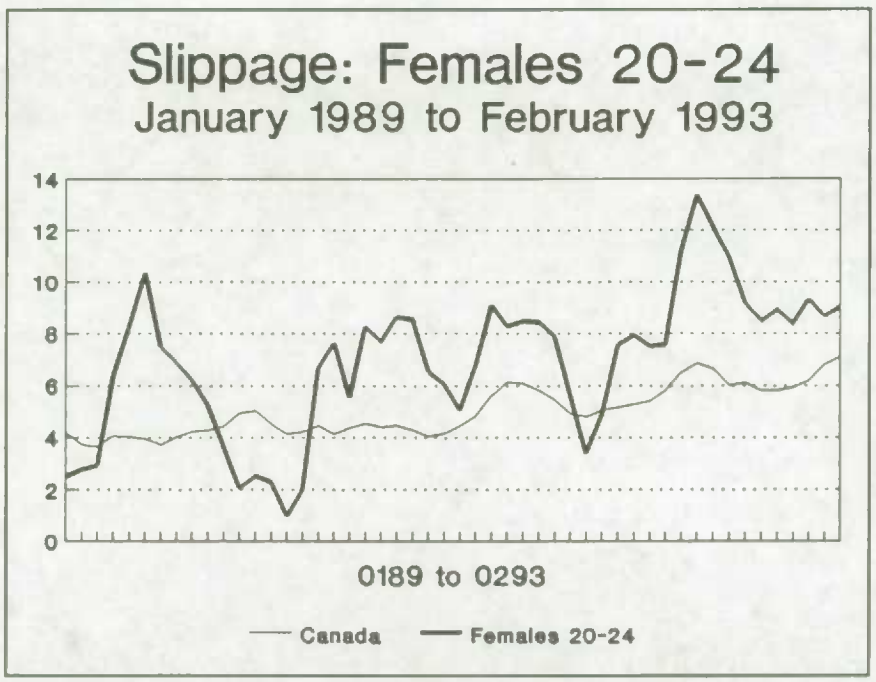




4.2 AGE-SEX TRENDS

The 20-24 year old age group is most prone to undercoverage. Males in most age groups experience higher undercoverage than females. Undercoverage is estimated to be highest for young males aged 20-24. In January 1993 it was 15% representing 146,100 individuals in a population of 974,000. The rates for females of the same age, though lower than the male rates, have been consistently above the national level since early 1992. The 20-24 age group tend to be most mobile. This makes them more susceptible to the types of undercoverage discussed in Section 7, Listing Individuals. Graphs of slippage rates for both males and females in the 20-24 age group along with the associated national series follow.


Slippage: Males 20-24
January 1989 to February 1993
14
12
10
8
6
4
2
0189 to 0293
Canada
Males 20-24


Slippage: Females 20-24
January 1989 to February 1993
14
12
10
8
6
4
2
0
0189 to 0293
Canada
Females 20-24

5. FACTORS INFLUENCING SLIPPAGE

Undercoverage is caused when either households or individuals within households that should have been included in the survey are missed. This can occur when:

Missing Households
Selecting From the Area Frame
Listing Dwellings
Classifying Dwellings
Missing Individuals Within Households
Listing Household Members

Clusters are selected from the area frame with probability proportional to the number of private occupied dwellings in the last Census upon which the survey design is based (Example: 1981 Census for March 1993). These size measures become outdated over time and contribute to both the level of slippage and changes in slippage. The following section gives more detail and an example.

There are many reasons why an interviewer could miss dwellings while listing a cluster. Maps may not be clear or areas may be difficult to map. This is particulary so in some rural areas. There are not always obvious physical indictions as to what is intended as a boundary. Re-listing exercises (Switzer 1987) showed that error in listing was minimal and, hence, not a major impact on the changes in slippage over time. It is, rather, a small and consistent influence perhaps with some seasonality.

Interviewers may erroneously declare a dwelling out-of-scope. The individuals in these dwellings would contribute to undercoverage. The most common situation is misclassifying a dwelling as vacant. Data from monitoring interviewers (Vacancy Check Program) indicates that between 2% and 5% of dwellings declared vacant actually contain respondents. These figures are not based on a representative sample of all interviewers and, hence, can only provide a gross indication. Another study (Switzer 1987) done prior to the improvements to the 1987 LFS Interviewers Manual indicated that one percentage point of slippage came from misclassified dwellings.

Once an interviewer has established contact with a dwelling, the members of the household are listed by applying the concept of "Usual Place of Residence." Interviewers may miss individuals either because this concept has been misunderstood by the respondent or the interviewer has incorrectly applied the concept. The respondent may also intentionally withhold information which would lead to missing household members. The contribution to slippage of missing individuals within households is further discussed in Section 7.

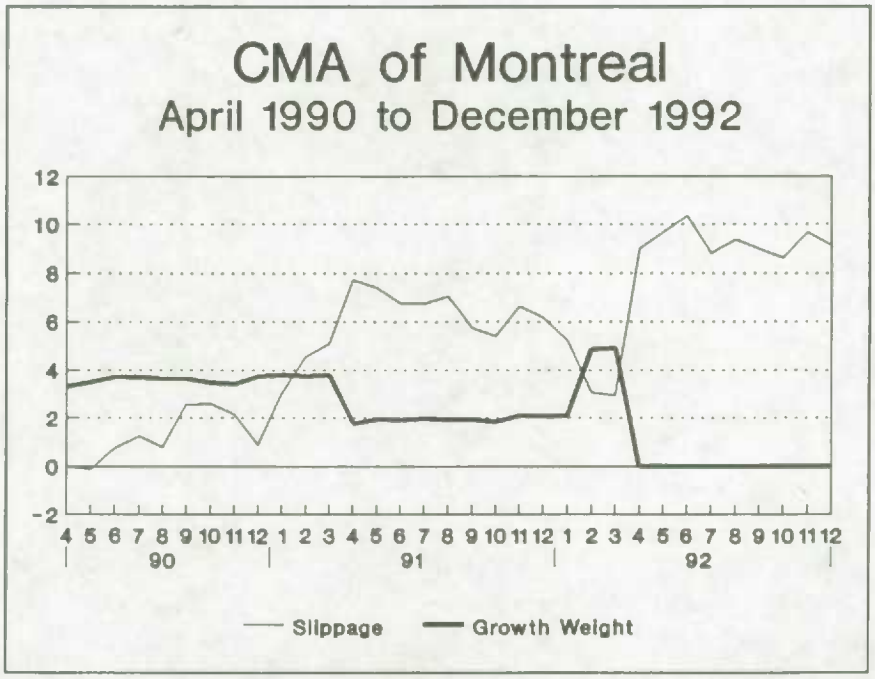
Differences between the post-censal estimates used in calculating the slippage rates and the "true" target population on the survey reference day can also contribute to movement in the slippage series. Post-censal estimates are based on counts from the past Census and administrative data to identify movements in and out of the Census population. Thus, they are subject to error. These errors misrepresent the level of undercoverage. Section 8 describes the post-censal estimation methodology and gives some measures of error.

6. SELECTION FROM THE LFS AREA FRAME

Clusters and PSUs are, as already described, selected with size measures that become outdated over time. The result is an increasing sampling variability of the design-based estimates. This could explain the increasing magnitude of month-to-month changes in slippage. In some clusters, for example, there may be many more dwellings in February 1993 than there were in June 1981. This is especially so on the rural fringe of cities where growth tends to concentrate in the form of new housing development. The hit or miss nature of encountering such growth using PPS size measures which do not reflect the growth can cause a significant movement in slippage.

Growth since 1981 is of no concern (to the sample design) if within each strata the growth rate is uniform across all areal units (ignoring any deterioration in the optimal nature of the stratification). Correlation coefficients between 1981 and 1986 census enumeration area counts within each strata demonstrate that growth does not occur in this manner. In cases where the stratum growth was greater than 30% between 1981 and 1986, the correlation was seen to take a wide range of values and was negative in some cases of extreme growth.

In some cases the number of dwellings listed just before a cluster rotates into the sample is so large that a cluster is sub-sampled in order to control survey costs ("growth" clusters). An extra weighting factor is applied to compensate for this sub-sampling. By examining the relationship between the slippage rate and the weights contributed by these clusters (expressed as a percentage of the post-censal estimate) it is seen that at times, slippage behaves as a reflection of the weight due to sub-sampled clusters. This implies a direct relationship between rotation of growth clusters and changes in the slippage rate. To illustrate this clearly we use the Census Metropolitan Area (CMA) of Montreal. Montreal is chosen because it is an obvious illustration and is not to be taken as representative of this problem.



In examining the graph note the period January 1992 to December 1992. The slippage rate is a virtual reflection of the cluster weights. In particular, an extreme monthly shift is observed between March and April 1992. In March the slippage rate was 2.9%. In April 1992 the rate climbed to 9.0%. One cluster that was in the sample in March and was replaced by a different cluster in April can explain most of this shift. The cluster in the sample in March had an expected yield of 86 households while the cluster which replaced it had a yield of 6. The design weight was 666. In this extreme example one cluster rotation explained a change in population of over 100,000.

Some clusters containing growth are not sub-sampled and, hence, are not represented in the graph. Closer investigation reveals that the large increase in slippage over the period December 1990 to February 1991 can be explained by the rotation of such clusters.

The conclusion is that most of the large fluctuations in slippage over time are due to the rotation of growth clusters. The designers of the survey chose the Rao, Hartley, Cochran method of sample selection in the SR frame in order to carry out frame updating (Drew et al. 1978) after the 1986 census. Unfortunately cost considerations prevented a frame update from occurring. The survey is presently undergoing a complete redesign, with new sample being introduced in October 1994, which will address this problem fully.

7. LISTING INDIVIDUALS

There are some difficulties in determining household members as the complex LFS questionnaire and instructions to the interviewer suggest. This was confirmed in a 1992 study to review and evaluate the LFS questionnaire (Price-Waterhouse 1992). It was found that the first question used to establish a list of household members is not easily understood and that the other two questions are perceived to be redundant. Some respondents found the concept of usual place of residence ambiguous while others found it easy to interpret and understand. It is likely that complex cases such as the individual moving around to find work or students are being missed. These cases may be more common among young males thus explaining higher undercoverage for young males. Some interviewers may not pursue household membership when they suspect they are missing individuals in order to reduce the risk of losing the cooperation of the respondent.

Respondents can also contribute to undercoverage by consciously withholding information. They may be involved in illegal activities or be involved in activities that they perceive to be unacceptable in their society. In some cases, the interviewer could be facing a cultural barrier to government officials. Other respondents may be reluctant to provide information which they perceive to be contradictory to information already provided to other officials. Individuals who receive social benefits, for example, may be reluctant to list everyone who usually resides in the household for fear of reducing their benefits.

The concept of usual place of residence may be more difficult to apply today than ever before because of the change in North American society in normative living arrangements. Perhaps in redesigning the survey, definitions and procedures should be revisited in view of this societal shift.

Another possible explanation for the high undercoverage of young males and, increasing, of young females is that they are insufficiently covered by the sample design when their usual place of residence is not their parental home and they reside in locations covered by the Special Area Frame such as remote work camps or institutions such as student residences.

8. POST-CENSAL ESTIMATES OF THE LFS TARGET POPULATION

Estimates of the LFS target population from the survey are compared against independent post-censal estimates to measure coverage error. The same post-censal estimates are used to adjust survey estimates in order to reduce coverage bias. Post-censal estimates are based on a complex combination of counts from the last census, administrative data, projection models, and error correction factors. Thus, they are subject to error, which in turn can misrepresents the problem of survey coverage error in the LFS. This section gives an overview of the post-censal estimation methodology and indicates the areas most susceptible to error.

Estimation Methodology

The methodology employed to produce estimates of the LFS target population is based largely on the methodology applied by Statistics Canada to produce official annual post-censal estimates of the census population. Extensive documentation of the latter appears in Statistics Canada (1987) while the former is in Bender (1992). The following summary borrows heavily from these two sources. It is presented to give the reader an overall flavour and, thus, does not do justice to the complexity of demographic techniques employed.

The estimation methodology described below is used for producing estimates at the province level only. Subprovincial estimates are also produced using a slightly different methodology. Coverage error is monitored less at the subprovincial level because of the high variability of the survey estimates.

Estimates of the LFS target population for use, for example, in the February 1993 survey are derived in three stages:

1. **Census Population On June 1, 1992** - Produced from last available Census day counts (June 3, 1986) by applying estimates of the components of change in the population derived from administrative data.

2. **Census Population On Survey Reference Day February 20, 1993** - The same methodology is applied. This step is distinguished from the first because the estimates from the first step have broader use inside and outside of Statistics Canada.

3. **LFS Target Population on Survey Reference Day February 20, 1993** - Estimates of the unsampled population are subtracted.

Change since the last Census is based on aging those individuals counted in the last Census and estimating the number and demographic characteristics of individuals who entered or left the Census population. This "Components of Change" approach is based on the following equation:

$$P_{t+i} = P_t + (B_{(t,t+i)} - D_{(t,t+i)} + I_{(t,t+i)} - E_{(t,t+i)}) + N_{(t,t+i)}$$

where

P_{t+i} = Census Population in any province at time t+i (reference date)

P_t = Census Population in last Census at time t

$B_{(t,t+i)}$ = Number of births between time t and time t+i

$D_{(t,t+i)}$ = Number of deaths between time t and time t+i

$I_{(t,t+i)}$ = Number of immigrants entering the province from outside Canada between time t and time t+i

$E_{(t,t+i)}$ = Number of emigrants leaving the province and Canada between time t and time t+i

$N_{(t,t+i)}$ = Net number of interprovincial migrants between time t and time t+i

Each component in this equation is produced by sex and single year of age. The components of change are derived from the administrative data sources listed in the following table. This table also gives the time lags from the occurrence of the event to when the data are received by Statistics Canada. Information on overall levels is received within 1 to 4 months while the distribution by single year of age and sex is received a year later. An exception is data for estimating interprovincial migration where the distribution by sex and age group is received at the same time as the totals themselves. Estimates for the reference date are produced by applying past trends and growth assumptions as well as treatments unique to the nature of the data source.

Data Sources for Components of Change

Component	Source Data	Detail	Time Lag (Months)
Births, Deaths	Provincial Registers	Totals Age, Sex	4 18
Immigration	Canada Employment & Immigration	Totals Age, Sex	2-3 18
Emigration	1. Family Allowance 2. Revenue Canada income tax files 3. US Immigration and Naturalization Service	Totals Age, Sex	4 18
Interprovincial Migration	1. Family Allowance 2. Revenue Canada income tax files 3. Last Census	Totals Age Group, Sex Age	1 18 quinquennial

Preliminary as well as revised source data are used. The post-censal estimates are continually adjusted to reflect new source data. Thus, a portion of each post-censal estimate is an adjustment for past errors. This approach results in the post-censal estimates being consistent with the most recent data but containing a component that does not represent people. The magnitude of the error correction factor is controlled. It is a negligible contribution to slippage.

The last stage in producing post-censal estimates is to remove those populations in the Census that are out-of-scope for the LFS. Growth assumptions are applied to counts from the last Census of the population residing on Indian reserves, of inmates of institutions, and of Canadians living abroad. Thus, the counts used in the estimates are based on data that are 1 to 6 years old. Counts of the number of full-time members of the Canadian Armed Forces are received monthly by location, year of age and sex from the Department of National Defence.

The following table gives the magnitude of each component of the post-censal estimate of the LFS target population aged 15+ for Canada and for Quebec for February 1993. The number of births is included but is not an additive component to the 15+ population. A province is included to give the magnitude of the internal migration component.

Post-Censal Estimate of LFS Target Population 15+
February 20, 1993

	Contribution (+ or -)	Canada (1000s)	Quebec (1000s)
Post-Censal Estimate of Census Population 15+ on June 1, 1992		21,668	5,539
Age To 15+	+	264	69
Births	+	293	71
Deaths	-	140	35
Immigration	+	149	25
Emigration	-	24	3
Internal Migration	+		-11
Error Adjustment	+	1	.5
Estimate of Census Population on February 20, 1993		21,919	5,584
Inmates of Institutions	-	338	92
Population on Indian Reserves	-	163	23
Armed Personnel: Internal External	 - -	 75 5	 10 1
Canadians Abroad	-	11	3
Total Not In LFS Target Population	-	592	129
Estimate of LFS Target Population		21,327	5,455

Limitations of Post-Censal Estimates

In general, estimates for births, deaths, and immigration are considered quite accurate. Estimates for emigration and internal migration are estimated indirectly and are, hence, subject to greater error.

Time lags in using the administrative data are one limitation of using post-censal estimates to represent the LFS target population. These lags are greater, by about a year, for distributions by year of age and sex than for the levels themselves. The estimation methodology assumes that patterns seen at the reference date of the administrative data apply at the survey reference date 1 to 18 months later.

Another limitation is that administrative data are a secondary data source. They reflect an administrative process rather than one designed exclusively to collect statistical information. There may be differences in coverage and concepts. An example of this is the use of the Revenue Canada personal income tax file to

estimate internal migration. The residence used from the tax file is the residence of the tax filer as of the date of filing rather than the residence on each of the monthly LFS reference dates. Also, there is a tendency for young females to be counted twice because they file themselves while being claimed as a dependent by another filer. Adjustments are made to account for these deficiencies.

The coverage of administrative data results in assumptions in the estimation methodology of varying credibility. The use of Revenue Canada's Family Allowance file and personal income tax file to estimate migration, for example, results in two assumptions regarding migration patterns worthy of questioning. First, the assumption is made that adults migrate with their children. The second assumption is that adults with children have the same migration patterns as those without. Again, adjustments are made to account for these deficiencies. Regardless, these assumptions are two reasons that migration estimates are considered less accurate. Their impact on slippage merits some investigation.

One implication of this discussion is that there is a component of slippage which can be viewed as observing shifts from the assumptions made in producing post-censal estimates. The magnitude and direction of these shifts influence slippage.

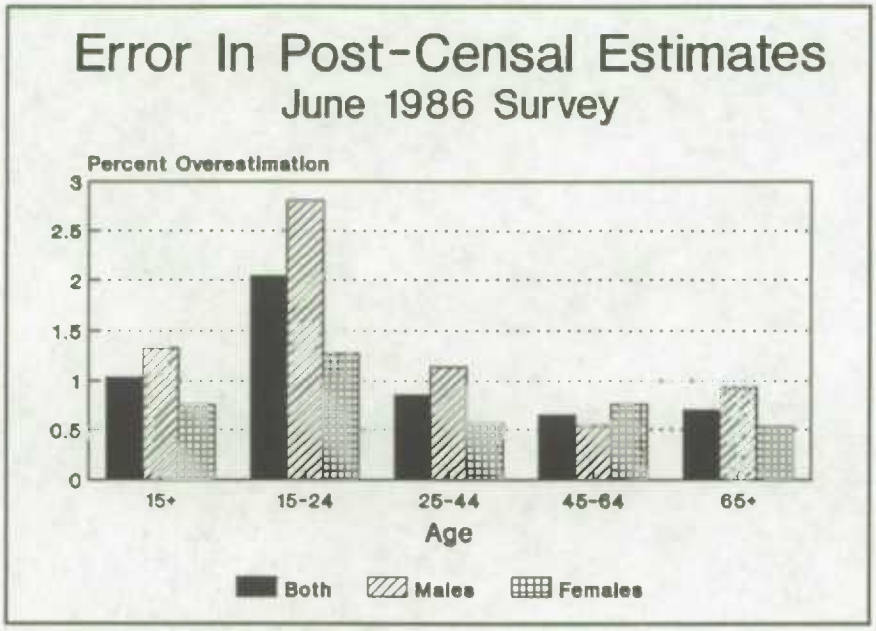
Evaluation

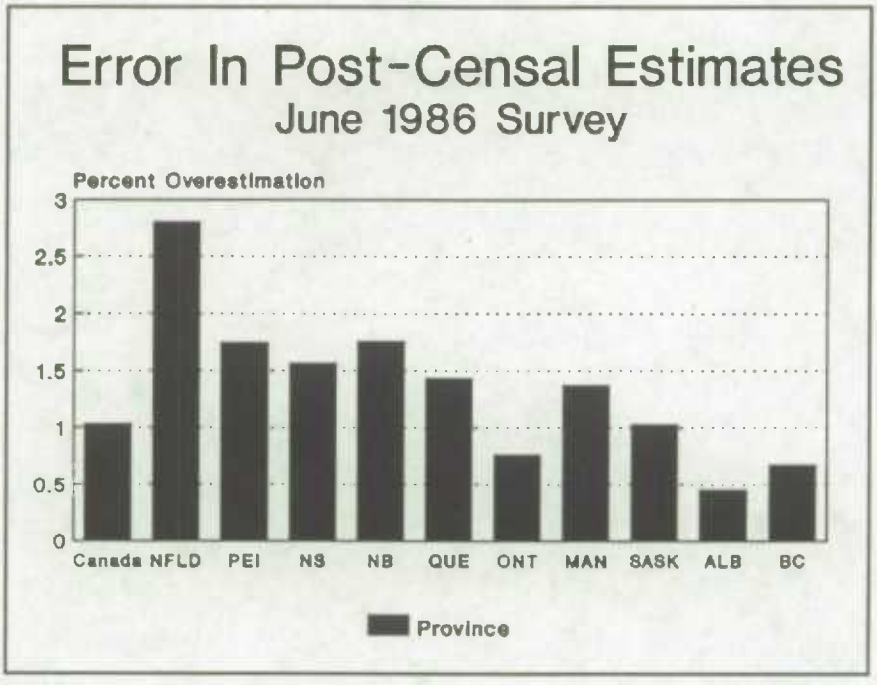
There is limited opportunity to evaluate post-censal estimates against an independent source. A comparison with the Census, however, is considered a reliable benchmark. The difference between the two, called the "error of closure," provides a measure of accuracy. The following table gives the error of closure, for 1981, 1986, and 1991, for the post-censal estimate of the LFS target population. They are presented as the percent by which the population on Census day was overestimated. The table also gives the impact on slippage of this error by comparing the reported slippage with what it would have been if counts on the Census day, a few days earlier, had been used.

Evaluation Of Post-Censal Estimates

	1991	1986	1981
LFS Population (15+) on Census Day (1000s)	20,789	19,388	18,367
Estimates Based on Previous Census (1000s)	20,740	19,588	18,287
Difference % Overestimated	-.2	1.0	-.4
Reported Slippage	6.1	5.6	4.7
Slippage Based on Census Day Counts	5.9	4.6	5.1

This table indicates that the error of closure (overestimation) was highest in 1986 at 1%. The following graphs show how this error is distributed by age and sex and by province. The largest error was 2.8% in estimating young males aged 15 to 24. The largest provincial error was 2.8% in Newfoundland. Error was generally lower in the Western provinces.





These comparisons with the Census show that error in the post-censal estimates of the LFS target population is estimated to range from -0.4% (underestimation) to 1.0% (overestimation). This magnitude of error, although within acceptable levels for a post-censal estimation methodology of this type, is a significant

component of slippage since it results in an -8% (underestimation) to an 18% (overestimation) error in slippage.

The above errors include errors in Census coverage as well as errors in the components of population change derived from administrative data. In January 1993 the Chief Statistician of Canada decided to adjust population estimates to account for Census coverage error. The post-censal estimates used in the LFS for adjusting survey estimates and for measuring coverage error will, therefore, be accordingly adjusted although the date of such a switch and the extent of historical revision are to be determined. Slippage will increase as a consequence of this change particularly for those individuals like young males who are undercovered the most by the Census.

9. CONCLUSIONS

This paper has reviewed the multitude of factors which can contribute to levels, changes, and longer term trends in survey undercoverage as measured for the LFS by the slippage rate. The mechanisms which cause movement in slippage have been reviewed and an attempt has been made, where possible, to provide quantitative information on the magnitude of the influence of each factor on slippage.

Slippage can be decomposed into three trends:

1. **Constant** - There is a group of factors which result in a constant influence, perhaps with some seasonality, on slippage. Coverage error from field operations fall in this category: missing dwellings in listing, misclassifying in-scope dwellings as out-of-scope, and missing individuals on the household roster because of conceptual difficulties regarding usual place of residence or respondents willingly withholding information. There may be a component of slippage that moves with the economy as individuals are more job mobile and, hence, more likely to be missed via usual place of residence.

2. **Oscillations of Increasing Magnitude** - The survey frame, via increasingly out-dated cluster and PSU size measures, contributes this pattern.

3. **Upward Trend** - The mechanism behind this trend, beyond recognizing that many of the assumptions in the survey design and in the post-censal estimation methodology are based on the 1981 and 1986 Census respectively, remains to be fully discovered.

This paper contributes to an understanding of coverage error in the LFS and an understanding of movements in slippage with the following lessons:

1. The survey design influences both the level of slippage and month-to-month change to an extent that had not been previously appreciated. In particular, the rotation of clusters in urban areas and PSUs in rural areas in which growth, relative to size measures used to design the survey (1981 Census size measures used from 1984 to 1994), has occurred can cause significant movement in slippage. Growth is reflected in the post-censal estimates less sporadically than in the survey estimates.

2. The post-censal estimation methodology has sufficient potential for error in estimating the LFS target population to account for a significant component of slippage particularly at the provincial level. The error of closure for post-censal estimates is considered acceptable for adjusting survey estimates.

3. Slippage is an indicator of more than coverage error. It reflects the currency of the frame as well as, to a lesser degree, the shift from assumptions used in producing post-censal estimates.

4. The following field operations exhibit a constant impact with, perhaps, some seasonality: missing dwellings in listing, misclassifying in-scope dwellings as out-of-scope, missing individuals on the household roster because of conceptual difficulties regarding usual place of residence or respondents willingly withholding information.

5. The highest potential for missing individuals in field operations occurs in compiling a household roster. Conceptual difficulties in understanding Usual Place of Residence have been reported by both the interviewer and the respondent in focus groups and interviews. Further, interviewers may not pursue the issue of household members in order to minimize the risk of losing the interview. A related issue is coverage of individuals living in remote work camps or institutional environments such as university residences.

6. Survey undercoverage is highest among young males aged 20-24 and, increasingly, among young females of the same age. While the factors discussed in item 5 are especially pertinent to these populations undercoverage of youth and males in particular seems to be the nature of household surveys. Census undercoverage rates in Canada and the US reflect this trend.

10. RECOMMENDATIONS

Survey Design

1. Improve monthly diagnostics to isolate the impact of cluster rotation.
2. Consider alternative treatments of growth that would minimize large fluctuations in design-based estimates due to the rotation of clusters which contain growth.
3. Investigate the coverage of individuals living in remote work camps or institutional environments such as university residences to gain insight into undercoverage of young males.

Questionnaire Design

1. Redesign the questionnaire and interviewer procedures regarding the determination of household members based on usual place of residence.

Field Operations

1. Introduce programs that provide regular measures of the contribution to slippage of field operations (listing dwellings in error, misclassifying dwellings as out-of-scope, listing household members). These measures should be probability-based and applicable to all interviewers.
2. Provider refresher training to all levels of field operations staff.

Post-Censal Estimates

1. Quantify the difference in month-to-month change between post-censal estimates and design-based estimates to see when the two methodologies show growth.
2. Continue evaluating the post-censal estimation methodology for interprovincial migration to determine the magnitude of the error in provincial post-censal estimates

REFERENCES

BENDER, ROSEMARY (1992), "Population Estimation Programme of Labour and Household Surveys Analysis Division," Statistics Canada internal document.

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, 4, 225-263.

LEMAÎTRE, G. E., and, DUFOUR, J. (1987), "An Integrated Method for Weighting Persons and Families," Survey Methodology, 13, 199-207.

PRICE WATERHOUSE (1991), The Review and Evaluation of the Questionnaires Used in the Labour Force Survey, Final report for Statistics Canada.

RAO, J.N.K, HARTLEY, H.O., and COCHRAN, W.G. (1962), "A Simple Procedure of Unequal Probability Sampling without Replacement," Journal of the Royal Statistical Society, B, 24, 482-491.

SINGH, M.P., DREW, J.D., GAMBINO, J.G., and MAYDA, F. (1990), Methodology of the Canadian Labour Force Survey 1984-1990, Ottawa, Minister of Supply and Services, Catalogue 71-526.

STATISTICS CANADA (1987), Population Estimation Methods, Canada, Ottawa, Minister of Supply and Services, Catalogue 91-528E.

STATISTICS CANADA (1989), Labour Force Survey Interviewer's Manual, Internal document, Ref. No. LFS-100.

SWITZER, KAREN (1987), "Reducing and Monitoring Slippage," Statistics Canada internal document.

HA 806

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA

1010140255