

SAMPLE DESIGN OF THE MONTHLY RESTAURANTS,
CATERERS AND TAVERNS SURVEYM.A. Hidiroglou, R. Bennett, J. Eady and L. Maisonneuve¹

Statistics on sales of establishments classified as restaurants, caterers and taverns have been collected since 1951. The sample has not been updated for births since 1968 and as a result, it is not representative of the current universe. This paper reports on several methodological aspects of the redesign. The sampling unit, sample design, sample size and allocation, data collection methods, edits and imputations, accumulations and calculations, frame and sample maintenance are described. The new survey will reduce manual procedures wherever possible. Collection, editing, imputation, tabulation and updating procedures will be completely computerized. Data collection will be decentralized and will take place via telephone.

1. INTRODUCTION

Establishments classified under Standard Industrial Classification Code 886 have been surveyed since 1951 through a monthly sample survey. This sample does not completely cover all establishments under SIC 886 since taverns, take-out restaurants, caterers and refreshment stands are excluded. The sample has not been updated for births since 1968 and the sample size presently stands at approximately 300 businesses.

¹ M.A. Hidiroglou, Business Survey Methods Division, R. Bennett, Merchandising and Services Division, J. Eady, Systems Development Division and L. Maisonneuve, Regional Operations Division, all of Statistics Canada.

As a result, it is not a representative sample of the current universe of around 30,000 which encompasses all establishments within SIC 886. This sample size fails to satisfy the needs of the Canadian System of National Accounts and the industry itself.

The frame for SIC 886 was updated and surveyed in 1976, 1977 and 1978. This operation was done under the heading of Restaurants, Caterers and Taverns Survey (RCTS). The RCTS ascertained the kind of business (KOB) and total sales of the businesses surveyed. The 1978 RCTS results were used to determine the sample size needed to support reliability criteria, provided by Merchandising and Services Division for a new monthly sample survey. This sample was drawn in December 1979 from an updated 1979 RCTS frame and data collection commenced in January 1980. This new monthly survey will run in parallel with the old monthly for a period of one year at which time the latter will be dropped. This new monthly survey departs in quite a number of ways from the old monthly survey. The old monthly survey was completely manually operated, the manual operations ranging from mailout to tabulation at head office. This new monthly survey will reduce manual procedures wherever possible. Collection, editing, imputation, tabulation and updating procedures will be completely computerized. Data collection will be decentralized and will take place via telephone by the Regional Offices.

Information from this survey will be used by members of the industry to compare their individual growth with that of the Province and in the creation of marketing techniques. It will be used by Federal and Provincial governments in the development of fiscal policies (cost sharing agreements) and to measure and forecast growth in the industry.

Other major users include the Restaurant Association of Canada, universities and tourism associations and agencies which are also involved in the development of the many facets of the food and beverage industry.

This paper reports on several methodological aspects of the redesign and will describe the sampling unit, sample design, the sample size and allocation, the data collection methods, edit and imputations, accumulations and calculations, tabulations, frame and sample maintenance.

2. COVERAGE

Eating and drinking establishments covered in this survey include all known businesses with establishments classified to the Standard Industrial Classification code (SIC) 886. These are broken down into the following KOB classification:

<u>Type of Business</u>	<u>K.O.B. Code</u>
Licenced Restaurants	88601
Unlicenced Restaurants Drive-in-Restaurants	88602
Take-Out Restaurants Refreshment Stands	88603
Caterers	88604
Taverns, Bars and Night Clubs	88605

The survey does not cover eating and drinking places owned by and operated as an integral part of hotels, motels and other accommodation businesses. As well, eating and drinking places classified to non-commercial establishments such as Armed Forces messes, private clubs,

legion branches, or service clubs were excluded from the survey. Finally, all eating and drinking places operated by establishments classified to an industrial sector other than Service Trades (e.g. manufacturing or retail trade) are excluded, since data for these are included in publications produced by surveys covering these areas.

To facilitate the presentation, it is convenient first to define certain terms which will be used throughout the text. Statistics on businesses covered by the Merchandising and Services Division of Statistics Canada are normally presented under one of two standard concepts. One is the location concept, under which every physically separate place of business is classified to its own specific kind of business classification. The other is the establishment concept, whereby the classification is based on the smallest separate accounting entry capable of reporting all elements of basic industrial statistics. Under the establishment concept, the sales activities of two or more business locations, not all of which are necessarily in the same industrial sector or kind of business, may be measured. For the purposes of the Monthly Restaurants, Caterers and Taverns Survey (MRCTS), the establishment concept is used. As a result, the reported data may include business activities at more than one location and/or in more than one kind of business. For establishments which have more than one location, their sales will be reported at the province by kind of business. Hence, the sales for locations belonging to the same province and kind of business will be agglomerated. For SIC 886, Merchandising and Services Division has partitioned the universe into chain and non-chain organizations. An organization operating four or more trading locations under the same legal ownership at the establishment level is defined to be a chain organization. All other organizations not meeting the chain definition are classified as independent organizations.

Total Net Sales and Receipts includes receipts from the sales of meals and lunches, alcoholic beverages and other merchandise. Excluded from these figures are sales taxes collected by businesses for remittance to any government agency and non-operating income such as service and carrying charges on credit accounts, bank interest and interest on investments, rent (Real Estate only) etc.

3. FRAME

The basic strategy for creating the 1979 Restaurants, Caterers and Taverns survey frame was to update the 1978 frame with information provided by the Business Register and the Retail Trade Survey. The Business Register provided businesses coded to SIC 886 having employees. The Business Register which receives its information through Revenue Canada, keeps a record of every business having employees and for whom payroll deductions are made. The Business Register is updated monthly for new businesses, firms going out of business and any name, address or classification changes. The Retail Trade Survey provided businesses having no employees for whom tax deductions are made. These businesses are obtained by using a supplementary area sample list. The area sample list was obtained by selecting a number of areas from a list covering all the regions in Canada with the exception of the Yukon and Northwest Territories, as defined by the Labour Force Survey. Following the 1966 Census of Population, the selected areas were then completely enumerated by field representatives and an area list of all business locations was created. This list was then matched against the main components of the survey frame. Any establishment duplicated on the main file by either a single or multi-location establishment was removed from the area sample list. Every month since the original enumeration, one twelfth of the selected areas has been completely re-enumerated in order to locate any new firms which may have come into existence

during the preceding year. The establishments remaining on the area list represented not only themselves but also similar establishments in the areas not selected for enumeration. A sampling weight, equal to the inverse of the probability that an area containing a particular establishment was selected, is associated with that establishment. The 1979 RCTS frame was therefore made up of units on the 1978 frame which were still active in December 1979, some 200 units from the area component provided by the Retail Trade Survey and some 6000 new establishments provided by the Business Register. These establishments were introduced into the Business Register during the course of the 1979 year. These new establishments were assigned an imputed sales value determined at the province and kind of business level and based on sales values reported for the 1978 RCTS. The sample for the monthly Restaurants, Caterers and Taverns Survey was then drawn from the 1979 RCTS frame. The sample size determination, allocation and selection are described in the section that follows.

4. SAMPLE SIZE DETERMINATION, ALLOCATION AND SELECTION

As was mentioned earlier, the 1979 RCTS frame is made up of chain and independent organizations stratified by province and kind of business. The independents are derived from the Business Register and the Retail Trade Survey. Those independents provided by the Retail Trade Survey are themselves a sample of establishments with no employees. The independents which are provided by the Retail Trade Survey will be termed as area units. These area units and the chain organizations are automatically included in the MRCTS sample. The remaining independents are further substratified within their province and kind of business classification into a take-all and take-some substratum. The overall sample size and allocation to these substrata is done at the province and kind of business level by taking into account precision constraints. These precision constraints are

defined in terms of coefficients of variation at the provincial and kind of business level. The sample size from the MRCTS was determined by using sampling rates derived from the 1978 RCTS frame. The 1978 RCTS was basically a mail survey. A sample of units which had not responded to the mail questionnaire was surveyed by telephone at the Regional Offices. Units which indicated that they did not wish to respond to the survey as a result of either receiving the mail questionnaire or being followed up by the Regional Offices had their sales imputed. This imputation was done at the provincial by kind of business by type of unit (independents originating from the area file or independents originating from the Business Register) level. All the chains responded to the survey and, therefore, none of their sales were imputed. Units which responded to the mail questionnaire were assigned a weight of one as were units which at that time indicated that they did not wish to provide data. Units which responded to the Regional Office follow-up and those that declined to respond to the follow-up were assigned a weight inversely proportional to their selection probabilities. In order to facilitate the development for sample size determination, some notation is defined. Let R_1 be the set consisting of all responding units to the survey, including those that were imputed. Units belonging to this set are used to compute totals. Let R_2 be the set consisting of all responding units to the survey which were not imputed. Units belonging to this set are used to compute sampling variances.

Let

y_{gpkj} = the sales associated with the j^{th} reporting unit in the g^{th} group, p^{th} province and k^{th} KOB,

w_{gpkj} = the weight associated with the above unit,

Q_{gpk} = the number of units in the g^{th} group, p^{th} province and k^{th} KOB which belong to R_1 ,

M_{gpk} = the number of units in the g^{th} group, p^{th} province and k^{th} KOB which belong to R_2 .

Note that the value for y_{gpkj} is zero for units which did not respond either to the mail or to the field follow-up. The weight w_{gpkj} is greater or equal to one for units which were followed up by Regional Office or which belonged to the area universe. For all other units, this weight is equal to one. p ranges from 1 to 12, denoting the ten provinces and the two territories. k ranges from 1 to 5, denoting the five previously defined kind of businesses. g ranges from 1 to 3, denoting the groups. These groups are chains [1], independents originating from the Business Register [2] and area units originating from the Retail Trade Survey.

The next step is to describe how the sample size and allocation are determined so as to satisfy simultaneously the provincial and kind of business requirements. Let $c(p,.)$ and $c(.,k)$ denote the required coefficients of variation for the p^{th} province and k^{th} KOB respectively. The marginal provincial and kind of business totals are:

$$\hat{Y}_{\cdot p \cdot \cdot} = \sum_{g=1}^3 \sum_{k=1}^5 \sum_{j=1}^{Q_{gpk}} y_{gpkj} w_{gpkj} ; p=1, 2, \dots, 12;$$

and

$$\hat{Y}_{\cdot \cdot k} = \sum_{g=1}^3 \sum_{p=1}^{12} \sum_{j=1}^{Q_{gpk}} y_{gpkj} w_{gpkj} ; k=1, 2, \dots, 5.$$

These marginal totals are used to compute KOB within province and province within KOB coefficients of variation. These within coefficients of variation are computed in order that they be equal within the classification of interest. Hence,

$$cw(.,k) = [c(.,k) \hat{Y}_{..k}] / \left(\sum_{p=1}^{12} \hat{Y}_{.pk}^2 \right)^{\frac{1}{2}} ; k=1, 2, \dots, 5;$$

and

$$cw(p,.) = [c(p,.) \hat{Y}_{.p..}] / \left(\sum_{k=1}^5 \hat{Y}_{.pk}^2 \right)^{\frac{1}{2}} ; p=1, 2, \dots, 12;$$

where

$$\hat{Y}_{.pk} = \sum_{g=1}^3 \sum_{j=1}^Q y_{gpk} w_{gpk}$$

Using these within coefficients of variation, a compromise first-round coefficient of variation for the p^{th} province and k^{th} KOB is computed as

$$c_o(p,k) = \frac{1}{2} [cw(p,.) + cw(.,k)].$$

The first-round coefficients of variation are next revised to satisfy the simultaneous KOB and provincial requirements by using the following iterative formula :

$$c_r(p,k) = \frac{c_{r-1}(p,k) c(.,k) c(p,.) \hat{Y}_{..k} \hat{Y}_{.p..}}{\left\{ \sum_{p=1}^{12} (c_{r-1}(p,k) \hat{Y}_{.pk})^2 \right\}^{\frac{1}{2}} \left\{ \sum_{k=1}^5 (c_{r-1}(p,k) \hat{Y}_{.pk})^2 \right\}^{\frac{1}{2}}}$$

where $r=1, 2, \dots, 10$. The iterative process revises the coefficients of variation at the province by KOB level so that they approximate in the best possible way the marginal provincial and KOB coefficients of variation. In our experience, the $c_r(p,k)$ values stabilized within four to five iterations. Let $d(p,k)$ be the chosen $c_r(p,k)$ coefficient of variation by province - KOB. It is then revised to take the area file variability into account. The revised coefficient is

$$dm(p,k) = \frac{\{[\hat{Y}_{.pk} \cdot d(p,k)]^2 - v_{3pk}\}^{\frac{1}{2}}}{(\hat{Y}_{.pk} - \hat{Y}_{3pk})}$$

where

\hat{Y}_{3pk} = estimated area file total for the $(p,k)^{th}$ stratum,

and

v_{3pk} = estimated area file variance for the $(p,k)^{th}$ stratum.

For each KOB by province classification, the cutoff points for take-all take-some substrata along with the corresponding sample sizes are computed. Note that the $dm(p,k)$ coefficients of variation are used as input. The algorithm which was given in Hidioglou (1979) is now described in steps. The province and KOB subscripts will be dropped in order to facilitate the presentation.

- (a) Within province (p) and KOB (k) and for a given level of precision dm , determine the sample size needed if simple random sampling without replacement were used with no take-all for the non-chain units, as

$$n(0) = \frac{\hat{N}_2^2 s_{2,M}^2}{(dm \hat{Y})^2 + \hat{N}_2 s_{2,M}^2}$$

where

\hat{N}_2 = estimated total number of units in the province - KOB
excluding chain and area file units,

$$= \sum_{j=1}^{Q_2} w_{2j} ,$$

\hat{Y} = total estimated population sales in the province - KOB
excluding area file units but including chains,

$$= \sum_{g=1}^2 \sum_{j=1}^{Q_g} w_{gj} y_{gj} ,$$

$s_{2,M}^2$ = estimated unweighted variance for the province-KOB,
computed over the M units which belong to R_2 and
which are not area or chain units.

$$= \frac{1}{M-1} \left\{ \sum_{g=1}^M y_{2,j}^2 - \frac{1}{M} \left(\sum_{j=1}^M y_{2,j} \right)^2 \right\}.$$

It has been assumed in the above variance formula that the y values are arranged in descending order. The subscript denotes that those units originally belong to the take-some substrata.

(b) Assuming that ℓ units had been included in the take-all substratum, the total number of units in sample for the $(p,k)^{th}$ stratum would be:

$$n(\ell) = \ell + \frac{(\hat{N}_2 - \ell)^2 s_{2, M-\ell}^2}{(dm \hat{Y})^2 + (\hat{N}_2 - \ell) s_{2, M-\ell}^2} .$$

Note that $s_{2, M-\ell}^2$ and $\bar{y}_{2, M-\ell}$, the respective variance and mean of the $(M - \ell)$ smallest independent units can be obtained recursively as:

$$s_{2, M-\ell}^2 = \frac{1}{(M-\ell-1)} \{ (M-\ell) s_{2, M-\ell+1}^2 + 1 - \frac{(M-\ell+1)}{M-\ell} (y_{2, \ell} - \bar{y}_{2, M-\ell+1})^2 \}$$

$$\bar{y}_{2, M-\ell} = \frac{(M-\ell+1) \bar{y}_{2, M-\ell+1} - y_{2, \ell}}{(M-\ell)} .$$

The boundary point between take-all and take-some is $y_{2, \ell}$.

The next step is to compute $n(\ell + 1)$. If $n(\ell + 1)$ is less than $n(\ell)$, the process is repeated till the inequality is reversed. That is

$$n(a-1) > n(a),$$

and

$$n(a) \leq n(a+1).$$

This inequality states that the iterative process is to be stopped. The number of units to be included in the take-all portion is "a" and the number of units to be included in the take-some is $n(a) - a$. The boundary point is then $y_{2, a}$ for the particular province and kind of business under study. The associated sampling rate is $f(a)$ where

$$f(a) = [n(a) - a] / (\hat{N}_2 - a).$$

Once all the boundary points and the associated take-some sampling rates had been determined for all the province by kind of business classifications, they were used as input for drawing the sample from the 1979 RCTS. Each province by kind of business classification (stratum) was substratified into chain, area, take-all, and take-some substrata. These substrata were created as follows. Establishments which were designated as chains were assigned to the chain substratum. Establishments which belonged to the area portion were assigned to the area substratum. All other establishments, the independents, were assigned to the take-all substratum if their sales exceeded the prescribed cutoff value $y_{2,a}$; otherwise, they were assigned to the take-some substratum. The process for selecting the sample within these province by kind of business strata was as follows. All establishments belonging to the chain, area and take-all substrata were selected with certainty. The take-some substratum was sampled at the rate $f(a)$ using simple random sampling.

5. EDITS AND IMPUTATIONS

This survey will have the sales data collected at the eight regional offices by telephone. Preprinted data sheets will be used to record data provided by firms participating in the MRCTS. Each data sheet will have spaces to record up to three sales figures for firms having a multiple kind of business or having operations in more than one province. Supplementary data sheets are available if more than three spaces are required. The information that is on each preprinted sheet is the Business Register Identification (BRID) number, the name and address, telephone number, province and kind of business and the associated number of locations. If any of this information is not correct, revisions will be made on the data sheet. The collected information will then be captured at the regional office and subject to some simple edits which will have been programmed for the mini-

computer. Records that do not pass these simple edits will be rejected. The operator will then have to pass back the questionnaire to the interviewer for correction. Those edits will insure that numeric fields designated as such are numeric, that they are of the correct length and that they either satisfy a range or are equal to some prespecified values. Records which have passed these simple edits are then transmitted to head office and transformed into transaction records. The Regional Office then acts as an emitter of data and head office as a receptor of data. The information on these transaction records is passed on to the sample file by matching on BRID number. Regional Office interviewers will also send a status code for each sales value obtained. These status codes which are given in Appendix A are used to categorize the response status of in-sample units and to identify units which must be edited or imputed.

It should be noted that most units are expected to report their sales data on a monthly basis. There are, however, some units which can only report on an annual, quarterly or even a thirteen periods basis. For units which can only report on a yearly basis, the monthly imputed value will be the previous year's sales values divided by the number of months in operation. The same type of imputation will hold for units which can only report on a quarterly basis. The imputation required for units which can only report on a thirteen period basis is more complex. Each of the thirteen periods corresponds to approximately twenty-eight days. The opening and closing dates for the thirteen periods vary from business to business. Hence, the opening and closing dates for one of these periods could be found either in the same month or in two consecutive months. Initially the sales associated with the thirteen period units are prorated by the ratio of the number of days in the month of interest to the number of days in the period which has a closing date in that month. These imputed sales, which are preliminary, are computed provided that the closing date is at least ten days

into the month. Otherwise, the sales are imputed using a ratio trend applied to the previous month's sales. Revised figures for the month are computed whenever sales are available for consecutive periods which have a closing date and opening date in that month. These revised sales are a function of the number of days in the month, the closing date and sales for the period which ends in that month and the opening date and sales for the next period which starts in that month. In most instances, the union of two consecutive periods will completely overlap one month and hence the associated revised sale will be a function of the data linked with those two periods. There are, however, some months of the year for which two consecutive periods have their closing date in that month and for which the next period has its opening date in that same month. For these cases, the revised sales are a function of the data associated with those three periods. An algorithm has been developed for the aforementioned cases and cases which can be more complicated (such as changing response status from one month to the next).

5.1 Editing the Monthly Data

The edits will be applied to records which have specific status codes chosen by the subject matter user. These status codes will be supplied by province, KOB and type of unit. This flexibility will allow for changes or expansion for the status code list as the survey progresses. A monthly trend edit will be computed by province, KOB and type of classification. This edit will be used to identify units whose current to previous period trend differs significantly from the general trend with the edited cell. The edited cell is that level of province, KOB and group aggregation for which edits are computed. For instance, before any collapsing occurs, the basic edit cell is the province by KOB and group classification, where the groups are chains, take-alls,

take-somes and area units. This collapsing occurs if there are not enough units within an edited cell to compute a stable mean and variance of the monthly trends. For a given group and kind of business, the collapsing is done across specified provinces. There are four levels of collapsing. The higher the level, the more provinces are included to participate. The choice of province combinations for collapsing purposes is based on similarity of annual sales means for those provinces, groups and kind of business classification. These annual sales means were computed using the 1978 RCTS results. Note that the collapsing will be done automatically if need be and that the level of collapsing will be determined by a preset minimum number of units needed to compute the means and variances. Note that some units will not be edited. These are births, deaths, non-respondents or those units becoming active or inactive on account of the seasonal nature of the operations. The notation defined below does not differ much from the one given in section 4. The symbol t is added to indicate that there now exists a time element (month and year).

$N_{gpk}(t)$ = number of units in the population at time t ,
in the g^{th} group, p^{th} province and k^{th} KOB,

$m_{gpk}(t)$ = number of units at time t which are used to
compute the trend, allowing for collapsing, for
the g^{th} group, p^{th} province and k^{th} KOB,

$n_{gpk}(t)$ = number of units in sample at time t , in the
 g^{th} group, p^{th} province and k^{th} KOB,

$n'_{gpk}(t)$ = number of unimputed units in $n_{gpk}(t)$,

$w_{gpkj}(t)$ = weight at time t associated with the j^{th} unit in the g^{th} group, p^{th} province and k^{th} KOB, (note that stratum jumper weights are automatically counted under this notation, see page 76).

$y_{gpkj}(t)$ = sales value at time t for the j^{th} unit in the g^{th} group, p^{th} province and k^{th} KOB,

$L_{gpkj}(t)$ = number of locations at time t for the j^{th} unit in the g^{th} group, p^{th} province and k^{th} KOB.

The ranges of g , p and k have been defined in the previous section. Note that j ranges from 1 to $n_{gpk}(t)$. The current month to previous month trends are computed as

$$r_{gpkj}(t) = \frac{y_{gpkj}(t) L_{gpkj}(t-1)}{y_{gpkj}(t-1) L_{gpkj}(t)} .$$

The mean and standard deviation of these ratios are used to construct intervals of tolerance. If the computed ratios lie within the intervals, they are said to have passed the ratio edit, otherwise they are flagged as a possible outlier. The span of these intervals is controlled by specifying a constant which determines how many standard deviations are tolerable on each side of the mean. Units declared as outliers will have their sales values imputed as described in section 5.2.

5.2 Imputation for Non-response and Outliers

Some of the imputations used in this survey have been mentioned earlier. These are imputing for number of locations, if need be, and prorating of sales figures for units which have a reporting period other than monthly. The imputation procedure operates by multiplying a non-responding unit's previous month's data by a measure of trend computed from responding units (excluding outliers) whose business characteristics are similar. Note that similar units may be obtained by collapsing if there are not enough data to compute trends. The collapsing will be across provinces by KOB and type of business.

During the initial month of the MRCTS, the averaged monthly sales (total yearly sales divided by twelve) from the 1979 Annual RCTS will be used as the base values to impute missing data. As the survey progresses, these imputed values should eventually be replaced by the monthly responses. The trend will initially be computed at the province, KOB and group level. The trend will be computed as

$$r_{gpk}(t) = \frac{\sum_{j=1}^{m_{gpk}(t-1)} y_{gpkj}(t) w_{gpkj}(t)}{\sum_{j=1}^{m_{gpk}(t-1)} y_{gpkj}(t-1) w_{gpkj}(t-1)} .$$

Note that the range is up to $m_{gpk}(t-1)$ for both numerator and denominator. The reason for this is that only pairs $y_{gpkj}(t)$, $y_{gpkj}(t-1)$ with valid information are entered into the numerator and denominator. Excluded from the trend computations are units which have not passed the edit test, or which have a zero sales value for month $t-1$. The number of units used in the computation of the ratio trend will be increased automatically by collapsing cells if there are not enough

data points. As with the editing collapsing, there are four levels to which cells may be collapsed. It is hoped that the nature of data collection for this survey will greatly lower the need for a large number of imputations. The non-responding units are then imputed by multiplying the above trend by its previous value. If no previous value exists, then the mean of the imputation cell will be used. The imputation cell is understood to mean a level at which the KOB and group cross-classification may be collapsed on the basis of province. Units declared as outliers will have their imputed value and their actual value printed out for review by subject-matter specialists. The appropriate Regional Office will then be asked to confirm the sales value for the unit declared as outlier. If the reported sales value is confirmed the reported value will be chosen. If the same unit is declared an outlier for several succeeding months, its weight will be changed to one and it will therefore become part of the take-all substratum.

6. ESTIMATION AND SAMPLING VARIANCES.

Once the data have been edited and necessary imputations carried out, the sample file is ready for aggregation. The basic building block for the MRCTS estimation process is the province, KOB and group type. Revisions of estimates are possible because Regional Operations Division are able to update data for the month previous to the reference month.

The notation is the same as that given in Section 5.1. The basic estimate of total for any province, KOB and group type classification is

$$\hat{Y}_{gpk}(t) = \sum_{j=1}^n y_{gpkj}(t) w_{gpkj}(t) .$$

Note that some of the y values may have been imputed. The associated variance will be given by

$$v(\hat{Y}_{gpk}(t)) = (1 - \bar{f}_{gpk}(t)) n'_{gpk}(t) s_{gpk}^2(t)$$

and $s_{gpk}^2(t)$ will not include imputed sales. This variance is computed as

$$s_{gpk}^2(t) = \frac{1}{n'_{gpk}(t) - 1} \left\{ \sum_{j=1}^{n'_{gpk}(t)} s_{gpkj}^2(t) - n'_{gpk}(t) \bar{z}_{gpk}^2(t) \right\}$$

where

$$z_{gpkj}(t) = y_{gpkj}(t) w_{gpkj}(t)$$

and

$$\bar{z}_{gpk}(t) = \frac{1}{n'_{gpk}(t)} \sum_{j=1}^{n'_{gpk}(t)} z_{gpkj}(t)$$

This variance allows for different weights within the same group type, province and KOB classification. Units within those classifications may have different weights on account of stratum jumpers. These stratum jumpers are units which change their kind of business during the course of the survey. They retain their original weight after changing KOB. The average correction factor is computed as

$$\bar{f}_{gpk}(t) = \left[\frac{1}{n_{gpk}(t)} \sum_{j=1}^{n_{gpk}(t)} w_{gpkj}(t) \right]^{-1} .$$

The variances associated with take-all or chain units are automatically zero. Once the estimation at the group type, province and KOB has been completed, estimation for higher levels of aggregation is additive over the proper sets. Furthermore, variance estimation for higher levels is also an additive operation which parallels the estimation process.

7. FRAME AND SAMPLE MAINTENANCE

The frame and the sample will be maintained with respect to births, deaths and amendments. This maintenance will be using as source the Business Register, the Retail Trade Survey area updating forms and information provided by the Regional Operations Division concerning the status of sampled establishments. The Business Register will provide births, deaths and amendments to the frame on a monthly basis. These updates will be for establishments which have employees. A sample of new establishments without employees will be provided on a monthly basis by the Retail Trade Survey. The Regional Operations Division will identify for head office establishments which are amended or are deceased. Amendments to sampled establishments can be changes in name, kind of business, Standard Industrial Classification or province. Deaths in the sample will be identified as establishments which have completely ceased their business operations. Amendments and deaths provided by the Regional Operations Division will be forwarded to the Business Register after the sample has been updated for these changes.

Updates originating from the Business Register are now discussed in greater detail since they will generate the bulk of changes to our frame and sample. These updates will be provided by the Current Update

File on monthly basis. This file is a record of all changes made to the Business Register in any operating cycle. It is essentially a file which contains a pair of records for every record that was added, updated or deleted in that cycle. The first record shows what was present before the change (blank, except for the Business Register Identification in the case of births) and the second record shows what was present after the change (blank, except for the Business Register Identification in the case of deletions). The Current Update File records will be put into three possible classifications. These are births, deaths and amendments. These changes will be applied to both the frame and the sample. Note that the Current Update File will be updated as well with information provided by the Regional Operations Division. These updates will be kind of business, Standard Industrial Classification, province or name and address changes. Amendments provided by the Current Update File refer to any change in information. The following changes will be considered as amendments: form of organization, survey status, activity status, registration status, kind of business, province, name and address, payroll deduction account number or telephone number. Amendments will be applied to establishments on the frame that match the Current Update File establishments. Births will be provided by the portion of the Current Update File which has a new Business Register Identification or a change of SIC to SIC 886. All establishments coded to SIC 886 will be included on the frame regardless of the SIC to which the head office may be classified. SIC 886 is a dynamic universe. Firms are constantly coming into existence, going out of business, merging with other businesses or splitting into two or more operations. This implies that these status changes should be very much kept up to date on both the sample and frame. It is for this reason that certain establishments should possibly be deactivated from the sample and master frame if they are found to be duplicates of incoming births. Finding these duplications is difficult for establishments on the frame which are not part of the sample. Deaths for these establishments will be

drawn to the attention of the Business Register at a later time. Note that it is only for establishments in the sample that the deaths can be quickly passed on to the Business Register and updated on the frame. Hence, for those establishments there will be little duplication on the frame.

The MRCTS is a survey which can be described as having a multiple frame, i.e. an area frame and a Business Register frame. To accommodate this multiple frame feature, weighting techniques such as the ones described in Hartley [1] could have been used. Since the area frame is small compared in terms of relative number of establishments with respect to the overall frame, the multiple frame problem will be dealt with by unduplicating the two frames. The unduplication of the frames will be implemented by matching the area sample units to the Business Register derived units on a monthly basis. Any business establishment in the area list which is matched to active Business Register records updated via the Current Update File will be deactivated from the frame and sample. The corresponding Business Register records will then be activated.

The formal process including births from the Current Update File records onto the frame and sample is several fold. Births provided by the Current Update File will first be examined for completeness of information such as name and address, province, regional office code and kind of business. These records which are judged to be in scope and complete will then be added to the frame. Births whose sales values exceed the cutoff value associated with the birth's province and kind of business will be automatically introduced into the sample. Units which have sales values lower than the cutoff will be sampled systematically at a rate prespecified for the particular province and kind of business combination. The establishments with no employees which are provided by the Retail Trade Survey will also be examined for the completeness of information. These area records which are complete and in scope will then be added to both the sample and frame.

8. CONCLUSIONS

The results of the Monthly Restaurants, Caterers and Taverns Survey will be published on a monthly basis. It is anticipated that preliminary estimates will be out six weeks following the month of reference. The final estimates will be published four weeks later. The tabulations will be at the Canada level by kind of business, by provinces, and by chains versus independents for Canada and the provinces. The estimated change from current year to previous year will also be published for the mentioned tabulation levels after sufficient time has elapsed to permit these calculations. Estimates of the coefficients of variation will be tabulated but will most likely be used for internal purposes.

There are several possible problem areas which will have to be closely monitored in order to ensure integrity of this survey. These are universe deterioration, response rates and measurement error. The universe deterioration is first discussed. SIC 886 is one of the most volatile SIC's. Restaurants constantly go into and out of business within a short time span. It is for this reason that it is important to have an updating procedure which keeps up with the changes. For every birth that comes into the survey, it is therefore important to identify, if possible, the name or owner under which the unit was previously known. Establishing such links, if they exist, will reduce duplication in the frame and slow down its deterioration. These links may be established by Nature of Business reports and by using the Merchandising Divisional Master File's updating procedures. The updating procedures contain programs which can establish possible links between births and units already on the file. Another source of deterioration for the frame will be changes of KOB for units not in the sample. Indeed, experience has shown that on the average there can be a 20% shift in KOB designation within one year. On a continuing basis all births will be contacted to obtain the kind of business before being placed on the frame.

Also, yearly sales attached with units not in sample will deteriorate over time on account of inflation. The effect will be that the cutoff boundaries established for the first year will have to be revised on the basis of sample results.

The response rate for this survey is anticipated to be high for a business survey, around 90%, as a result of the collection procedures utilized. Head Office will advise each Regional Office of the monthly response rate.

Another beneficial effect for field collection is that updates, be they name and address or changes of units into more or less KOB's, will be closely monitored for the sampled portion of the universe.

As Wolter et al [3] have reported, one potential source of bias for this type of survey will be non-sampling errors. Respondents who do not have book figures will provide an estimate of their sales. Awareness of response errors of that type can be brought to attention by keeping track of revisions to monthly sales.

ACKNOWLEDGEMENT

The authors are grateful to the referee for comments which have much improved this paper.

RESUME

Depuis 1951, on a cueilli des données sur les ventes des établissements qui sont classés comme restaurants, traiteurs et tavernes. La base d'échantillonnage n'a pas été mise à jour, en ce qui concerne les créations, depuis 1968; par conséquent elle ne représente pas bien la population actuelle. Cet article indique quelques concepts méthodologiques de la révision de cette base. L'unité d'échantillonnage, le plan de sondage, la taille et l'allocation de l'échantillon, les méthodes de la cueillette des données, la contrôle et l'imputation, les accumulations et les calculs, l'entretien de la base et de l'échantillon sont tous décrits. La nouvelle enquête réduira les opérations manuelles dans la mesure du possible. Les procédures de cueillette, de contrôle, d'imputation, de totalisation et de mise à jour seront entièrement informatisées. La cueillette des données sera décentralisée et sera faite par téléphone.

REFERENCES

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

- [2] Hidioglou, M.A. (1979), "On the Inclusion of Large Units in Simple Random Sampling", Proceedings of the Survey Research Methods Section, American Statistical Association.

- [3] Wolter, Kirk M.D., Isaki, Cary T., Sturdevant, Tyler R., Monsour, Nash J., and Mayes, Fred M. (1976), "Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys", Proceedings of the Business and Economic Statistics Section, American Statistical Association.

APPENDIX A

<u>Provincial Codes</u>		<u>Status Codes</u>	
10	Newfoundland	01	Acceptance
11	P.E.I.	02	Estimate by Respondent
12	Nova Scotia	03	Respondent Unable to Estimate
13	New Brunswick	04	Refusal
24	Quebec	05	Closed -Out-of- Business
35	Ontario	06	Increase due to Newly Acquired Liquor Licence
46	Manitoba	07	Increase/Decrease due to Special Events, Conventions, Seasonal Fluctuations, &c.
47	Saskatchewan	08	Increase/Decrease due to change in number of locations
48	Alberta	09	Decrease due to temporary closure (Renovations, Fire, &c) or Increase following Re-opening
59	British Columbia	10	Closed - Seasonal
60	Yukon	11	Out of Scope
61	N.W.T.		