

## **THE EVALUATION OF STATISTICAL PROGRAMS: A CASE STUDY**

**François Vaillancourt**  
**Université de Montréal**  
**Montréal, Quebec**

**Frank Clayton**  
**Clayton Research Associates**  
**Toronto, Ontario**

**Garry Sayant**  
**Statistics Canada**  
**Ottawa, Ontario**

**Abstract** — This article summarizes an evaluation of some Statistics Canada household surveys. The results have particular bearing on the usefulness of specific tools in assessing programs whose main objective is to produce data.

**Résumé** — Cet article est un court compte-rendu de l'expérience des auteurs au moment de leur récente évaluation de certaines enquêtes sur les ménages conduites par Statistique Canada. Il contribue en particulier à renforcer l'idée de l'utilité d'outils spécifiques lors de l'évaluation des programmes dont l'objet principal est la production de données.

**THE PURPOSE OF THIS** article is to summarize the experience of the authors in recently evaluating some household surveys of Statistics Canada. It adds, in particular, to knowledge on the usefulness of specific tools on assessing programs whose main objective is to produce data. The evaluation was carried out according to the usual practices of the Government of Canada as implemented in the Program Evaluation Division of Statistics Canada.

The article is divided into three parts. In the first, we present background information on the process of the evaluation and on existing issues in the area of the analysis of statistical programs. In the second, the effectiveness of various evaluation tools is discussed and in the third, the evaluation process and reports are examined.

## BACKGROUND INFORMATION

The evaluation procedures followed in evaluating the asset and debt, consumer finance, family expenditures, and facilities and equipment survey programs of Statistics Canada were as follows:

### *Evaluation assessment phase (July 1985–April 1986)*

- decision to evaluate the programs by Statistics Canada management, assignment of this task to a manager of the program evaluation division, and selection by this manager of outside experts
- identification of users to be interviewed and preparation of interview check list
- interviews
- preparation of evaluation assessment report identifying issues for the formal evaluation
- discussion of evaluation assessment report with program management (draft) and top management (final version) of Statistics Canada

In April 1986 the management of Statistics Canada decided that an evaluation should be carried out and set its terms of reference. As a result, Phase 2 was carried out.

### *Evaluation phase (May 1986–April 1987)*

Examination of information needs in the 1990s through:

- examination of interview notes
- review of the American literature
- examination of Canadian statistical trends
- one-day user seminar

Review of the existing survey procedures through:

- discussion with program management
- examination of program documentation

Assessment of various alternatives through:

- formulation of choice criteria
- detailed examinations of options

- discussion with program management

Preparation of recommendations presented to senior management

The first phase corresponds approximately to the first three phases suggested by Winberg (1986), and the second to the last three.

The evaluation took into account that the recent literature on statistical programs examines the consequences of budget cutbacks in statistical agencies on data availability (Pearson, 1984; Wallman, 1982) and, more generally, on society (Bonnen, 1983). That literature incorporates the assumption that statistical data are more (structural data such as income distributions and expenditure patterns) or less (market data such as prices) public goods. In general, authors (e.g. Sims, 1984; Spencer, 1984) agree that cost-benefit analysis, while providing a useful framework for evaluative purposes, is less useful in evaluating data programs because measurement of benefits is difficult. Indeed, Sims argues that cost-benefit models cannot be used to evaluate data programs, but evaluations should examine whether (a) the data-gathering activities are carried out correctly; (b) the desired data are being obtained and, if so, are being obtained efficiently; and (c) the data obtained are of the right type. He also points out that evaluations should not concentrate on easily measured costs and that groups and institutions that will likely make use of the data should be contacted to inquire into the nature of that use.

## THE EVALUATION TOOLS

The evaluation procedure generated conclusions concerning the usefulness of five data-gathering techniques: interviews with users, literature review, examination of statistical trends, user seminars, and examination of internal documentation.

### Interviews with Users

We conducted 31 interviews, 23 within the federal and provincial governments and the remainder within universities and the private sector. (See MacKay (1987) for another example of a list of interviewees and, more generally, of a program evaluation). Interviewees were selected from a list of recent purchasers of microdata files and special tabulations. The predominance of public-sector users is explained by the structural nature of the outputs examined: Very few private-sector users are interested, for example, in the level of poverty or the distribution of income. The user selection also

reflects consultants' prior knowledge regarding ongoing use of the data. Our experience leads us to make several observations:

1. The use of an incomplete sample of program users (users of publications were excluded because they were impossible to identify) did not cause a major problem for two reasons. First, the usefulness of users' comments on methodological and technical issues is a direct function of users' degree of sophistication in using the data. Thus, microdata file users are significantly more sophisticated than special tabulation users. Indeed, and second, the latter are in some cases no more sophisticated than publication users, treating special tabulations like published tables, using them in routine tasks (index calculations, etc.) and erring, at least by omission, in their use of the data.<sup>1</sup>
2. It is useful to interview a group of individuals, rather than a single person, within a given formal user (government department, for example) since use within a given formal user often varies by branch, division, and so on. Indeed in some cases our group interviews led to an exchange of information between individuals employed by the same user.
3. The more sophisticated users have often already tried, and sometimes succeeded, in influencing the data program. They can frequently produce existing written material documenting past requests. When preparing their evaluation, evaluators must be careful not to be overly influenced by such a sub-group of users.

## Literature Review

The purpose of the literature review was to identify trends in sociodemographic variables (one-person, divorced/separated, or blended households; aging population; working women; etc.) and expenditure behavior (electronic products; time-saving goods and services; financial services; etc.) that could influence the data needs of decision makers and thus the desired output of the data programs. This information was used in preparing the user seminar and part of the evaluation report. Our experience leads us to make the following comments:

1. The literature review was useful. It allowed us to identify trends (increase in work at home; increase in home health care) that were not raised by users, in interviews or at the seminar, and that our analysis of statistical trends using Canadian data could not pick up. This may be explained by the fact that most of the literature reviewed was for the United States, a

society where sociodemographic changes often precede the corresponding changes in Canada.

2. The most useful index in our literature search was the *Business Periodical Index*; the journal most frequently referred to was *American Demographics*. The future studies type of journal (*Futures Studies*, *Long Range Planning*, etc.) did not prove very useful in this case.

## Examination of Statistical Trends

The statistical trends evidenced by Canadian data were also examined in order to identify data needs. This information was used in preparing the user seminar and part of the evaluation report. We noted the following point:

1. In the case of Canada, the data available was more useful in identifying sociodemographic trends that were likely to affect expenditure patterns than in identifying changes in these patterns.

## User Seminar

The user seminar was attended by users from the private and public sectors and by program management. Its purpose was, as its title "The Household Surveys: Directions for the 1990s" indicates, to discuss with users the general trends in sociodemographic and economic variables, their impact on data needs, and specific changes to be made to the surveys. We made the following observations:

1. It is useful to send a small amount of material to participants in order to help them prepare for the seminar. In our case, they were sent the list of issues to be addressed in the evaluation, and a few tables on key trends.
2. It is stimulating to have users from both the private and the public sector attend such a seminar, as their perspectives on the usefulness of various surveys differ markedly.
3. In this case, seminar participants did not identify trends unknown to the evaluators, but were able to point out the trends' specific implications for various data needs.

## Examination of Internal Documentation

As part of the evaluation process, the evaluators met with program management to discuss the various issues raised in the evaluation assessment. As a

result, they were provided with internal documents on various aspects of the program. This experience thus leads us to note the following:

1. It is extremely important to have access to internal documents held by program management so as to ensure the adequacy of evaluators' recommendations. Ideally, this is done in a friendly and cooperative way (as was the case here), with the quality of the program being the concern of both management and evaluators. If such a relationship cannot be established (see Blais, 1986, for a theoretical discussion of the reasons for this), then the submission of a draft of the report to program management, which must either refute it with the appropriate documentation or accept it, may be an effective way of obtaining the required information.

## **THE EVALUATION PROCESS AND REPORTS**

Two reports were produced during the evaluation process: an evaluation assessment report (January 1986) and an evaluation report (February 1987).

The evaluation assessment comprised three main parts: a profile of the various surveys, a description of the concerns of users (survey coverage, reporting unit, etc.) identified by the interviews, and a recommendation on the issues to be addressed in an evaluation study. The second part directly reflects the data needs expressed by the users, with indications of how often (majority, a few, etc.) these were expressed. In preparing the third part, the evaluation team assessed the relative importance of these data needs and took them into account in arriving at the issues to be addressed in the evaluation study.

The evaluation report comprises four main parts. The first is a review of current and future information needs that provides a detailed description of the information being collected and identifies its main weaknesses (different reporting units, data overlaps, etc.), using, in part, the U.S. experience as a reference point. The second is a review of the existing surveys (sample, etc.) and of the issues that arise from this review (sample size, etc.). The third and core part of the report is an assessment of alternatives, which required the identification of criteria to be used in assessing the four options put forward by the evaluators. The fourth is a major appendix entitled "Sociodemographic Trends and Information Needs in the 1990s."

Both reports were submitted to the management of Statistics Canada. The conclusions of the first one were an input in the mandate that guided the formal evaluation. With respect to the evaluation process and the evaluation reports, the following noteworthy points can be made:

1. The selection of outside experts should be made with the objective of complementarity in mind. In this case, one was from the private sector, an expert on housing data, knowledgeable about expenditure data, and English speaking, and the other was from the university sector, an expert on income data, knowledgeable about expenditure data, and bilingual (English and French speaking).<sup>2</sup> As a result, users were interviewed in their language of preference, and expertise was available on all programs evaluated. More important, two different perspectives and two ways of approaching<sup>3</sup> the evaluation process were provided to Statistics Canada.
2. As Winberg (1986) suggests, the use of a two-step procedure with an assessment of the opportunity to carry out a full evaluation study once users' opinions have been collected is appropriate, as it facilitates the optimal allocation of scarce evaluative resources.
3. The recommendations of the evaluation report must rely on clearly identified criteria. In our case four major criteria were used:
  - user adequacy, which took into account three items:
    - the amount of information collected
    - the frequency
    - the geographical coverage
  - respondent burden, which was used to assess the feasibility of integrating surveys
  - survey administration, which took into account both contamination effects between surveys and bunching effects on staffing
  - cost

Here, cost was only one of the four criteria to be used in the evaluation.

## CONCLUSIONS

The main conclusions one can draw from this study are that (a) user interviews, even from an imperfect sample, are useful tools in evaluating data programs; (b) assessment of future needs can be enhanced through the use of such tools as literature reviews, user seminars, and analyses of statistical trends; (c) a step-wise evaluation process is preferable to a continuous one; and (d) criteria other than costs can be used to evaluate programs but must be developed for each case.

## REFERENCES

- Blais, R. (1986). Résistance et collaboration à l'évaluation de programme: Un cadre d'analyse et une modèle théorique. *The Canadian Journal of Program Evaluation, 1*, 47–58.
- Bonnen, J.T. (1983). Federal statistical coordination today: A disaster or disgrace. *American Statistician, 37*, 87–92.
- MacKay, S. (1987). The Manitoba air ambulance program: An evaluation assessment. *The Canadian Journal of Program Evaluation, 2*, 61–69.
- Pearson, R.W. (1984). The changing fortunes of the U.S. statistical system 1980–1985. *Review of Public Data Use, 12*, 245–269.
- Peskin, H.M., & David, F.L. (1984). Theory of an optimal data base. *Review of Public Data Use, 12*, 45–53.
- Sims, C.A. (1984). Can we measure the benefits of data programs? *1984 Proceedings Social Statistics Section, 17*, 60–67.
- Spencer, B.D. (1984). Towards conducting benefit cost-analysis of data programs. *1984 Proceedings Social Statistics Section, 17*, 46–51.
- Wallman, K.K. (1982). Federal statistics: The effect of program costs on availability, utility and quality. *Review of Public Data Use, 10*, 241–259.
- Winberg, A. (1986). A phased approach for conducting program evaluation. *The Canadian Journal of Program Evaluation, 1*, 1–11.

## NOTES

1. By, for example, comparing results of the expenditure survey for *all* Canadian households and for *urban* households, or by treating families (income survey) and spending units (expenditure survey) as identical.
2. Frank Clayton is president of Clayton Research Associates, a Toronto firm specializing in housing and retail trade studies; François Vaillancourt is a professor of economics who, in 1983–1985, was research coordinator, Income Distribution and Economic Security, for the Royal Commission on the Economic Union and Development Prospects for Canada.
3. Exaggerating somewhat, one could say that in obtaining information on a given point, Frank Clayton was more likely to seek out the relevant expert through the use of contacts, and François Vaillancourt was more likely to seek out the relevant paper through a literature search.