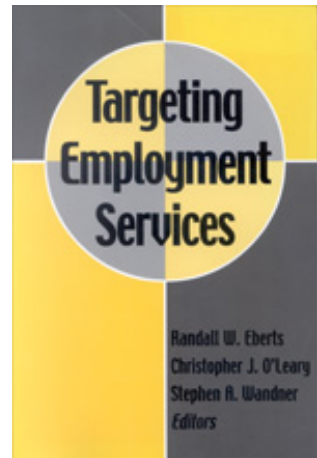

Upjohn Institute Press

Targeting Reemployment Services in Canada: The Service and Outcome Measurement System (SOMS) Experience

Terry Colpitts
Human Resource Development Canada

Jeffrey A. Smith
University of Maryland



Chapter 10 (pp. 283-308) in:

Targeting Employment Services

Randall W. Eberts, Christopher J. O'Leary, and Stephen A. Wandner, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2002

DOI: 10.17848/9781417524440.ch10

10

Targeting Reemployment Services in Canada

The Service and Outcome Measurement System (SOMS) Experience

Terry Colpitts
Human Resources Development Canada

The Service and Outcome Measurement System (SOMS) was developed by Human Resources Development Canada (HRDC) to be a tool for promoting employment. SOMS was intended to help frontline staff in local public employment service offices counsel job-seekers about the best strategies for gaining employment and to assist analysts and managers in determining the best employment and/or training strategies for specific client groups. A microcomputer-based prototype of SOMS was built in 1994. It had three main elements: 1) a relational database of client specific information for employment insurance beneficiaries and/or participants of HRDC employment or training programs, 2) a means for examining the results of past services provided by the public employment service, and 3) a computerized model to predict what services would most benefit a particular job-seeker. In 1997, an algorithm was added to SOMS for predicting what service would best promote employment among groups defined by geographic and demographic characteristics.

While SOMS has not been adopted in Canada, many useful lessons were learned in the course of its development and pilot testing. This chapter attempts to communicate the most important of those lessons while telling the story of SOMS. We begin by describing the policy context of SOMS. We then briefly explain the technical structure of SOMS, how SOMS could be used by frontline staff to assist job-

seekers, and how the model could be used to manage job-seeking by groups. The chapter concludes by reviewing some recent events in SOMS development, and reflecting on SOMS prospects for the future.

BACKGROUND

SOMS originated as a contribution by the Strategic Policy branch to the effort within HRDC known as the Knowledge Product Strategy.¹ SOMS built upon the multitude of evaluation studies performed by Strategic Policy's Evaluation and Data Development (EDD) branch during the prior 15 years. EDD viewed SOMS as a user-friendly vehicle for letting scientific research inform the management and practice of employment service delivery. Relying on an extensive client database summarizing past patterns of client services and outcomes, SOMS was intended to inform the choice of employment services for over four million annual customers of HRDC's nationwide network of local Canada Employment Centers (CEC).

Leading-edge evaluation techniques used within EDD formed the foundation for SOMS. However, to ensure that SOMS resulted in a user-friendly tool for management and practice, three development principles were established: 1) to link internal and external files to provide a detailed, sole source, multiple-year record of interventions provided to clients, their labor force participation and earnings history, as well as standard sociodemographic characteristics; 2) to develop and test statistical models to determine "point-in-time" intervention impacts at the client-specific level of detail; and 3) to incorporate the data and models in an interactive, micro-based system.

The SOMS prototype delivered to senior HRDC executives in the fall of 1994 was faithful to these principles as well as to the overriding objective of using research to inform practice. A series of SOMS demonstrations made to various groups in HRDC's national headquarters and many regional offices resulted in strong positive support for the SOMS initiative. There was so much support for the project and hopes were so high that SOMS developers tried to cool expectations.

SOMS was not intended to serve as an “expert system” to replace employment security officers, a potential trumpeted by some executives but feared by local office staff.

Concerns were also expressed about the privacy of client information held in the SOMS database, which was sometimes referred to as the “Big Brother” database. Some critics took the alternative position that the SOMS database was faulty, despite extensive data-checking and scrubbing routines employed by EDD. These criticisms and how they were addressed are explained in the following discussion of the four main system components and their historical development.²

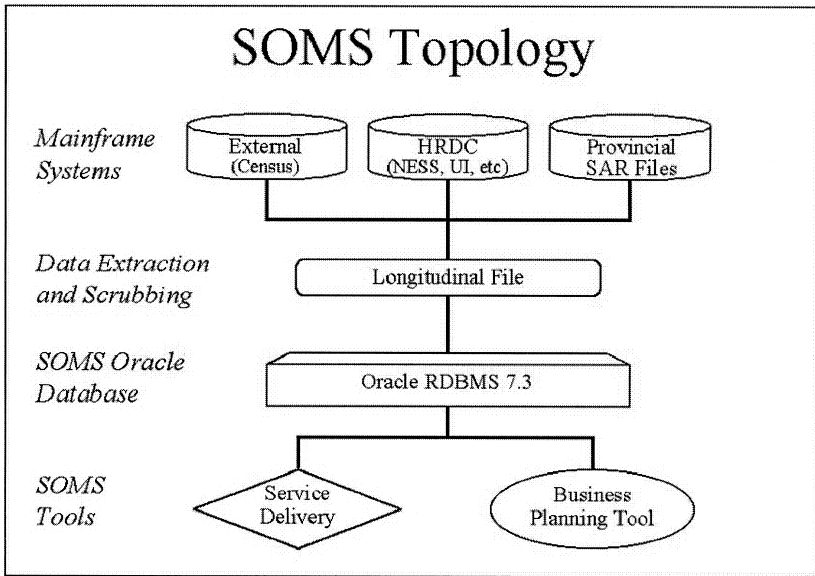
SOMS RELATIONAL DATABASE

The core of SOMS is a large relational database system. In the absence of highly reliable and credible data that can be accessed quickly, SOMS’s other components would not be acceptable to practitioners. The SOMS topology of data sources and preparation are summarized diagrammatically in Figure 10.1.

The initial step in the data-building process was extraction of information from 19 different administrative silos. Sources for these data included HRDC, provincial, and external mainframe systems. This compilation required 18 months and was completed by EDD staff in December 1995. The data is longitudinal in nature, meaning it contains information on individual clients over time. Nine years of data covering the period 1987–1995 were originally archived in SOMS. Programming specifications were defined for more than 2,000 variables grouped into four modules—individual, interventions, providers, and outcomes.

Extensive data-scrubbing routines were used in creating the longitudinal client database. In 1996, Oracle database software was selected as the HRDC standard for the regional database system, and by early 1997, an Oracle-based SOMS regional database system was operational. This database accommodated about 250 of the most important variables from the large longitudinal file on over 10 million clients.³

Figure 10.1 SOMS Topology

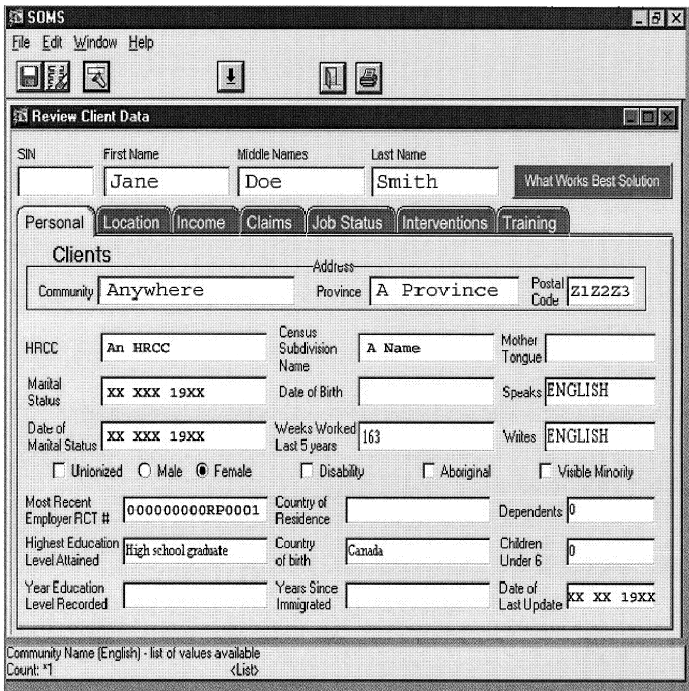


SOMS SERVICE DELIVERY TOOL

While the SOMS database was being constructed between 1994 and 1997, a prototype called Client Monitoring System (CMS)⁴ was being used for focus-group testing in 30 Human Resources Centre of Canada (HRCC) offices located in six main metropolitan areas. Figure 10.2 shows a graphical user interface screen from SOMS, similar to that used in the CMS prototype, which is used for reviewing client data.

CMS contained 6,000 records of HRDC clients who had been surveyed in 1994 as the first step in an evaluation of an initiative called the Employability Improvement Program. The focus group testing occurred during a sensitive period. HRDC had been formed only three months earlier by combining all or part of four previous federal departments. At the same time that its reorganization was under way, the federal government announced a workforce reduction of 25,000 full-time

Figure 10.2 SOMS Graphical User Interface to Review Client Data



NOTE: By entering a Social Insurance Number into this screen, the service delivery person obtains access to a rich data source on a client’s sociodemographic characteristics (Personal); data from the 1991 Census on the area in which they presently live (Location); income, earnings, unemployment insurance (UI), and social assistance benefits received over a multiple-year period (Income); detailed information on UI claims over a multiple-year period (Claims); a multiple-year record of employment, unemployment, and not in the labor force spells (Job Status); a multiple-year record of HRDC interventions provided (Interventions); and a multiple-year record of training provided (Training). Each of these information sections is shown as a file tab near the top of the record. The Personal tab (highlighted) is the one active in the screen above. At the far right and near the top, there is a button labeled “What Works Best Solution.” By pressing this button, it is possible to view which of about 25 possible HRCC interventions will lead to the best result for the client in terms of improving income, earnings, saving employment insurance, improving time employed, or reducing dependency on employment insurance. The solution is unique to the individual. The “What Works Best Solution” can be of assistance in making a service delivery decision but it is not a replacement for the good judgment of the counselor.

staff equivalents with HRDC's share of the reduction set at 5,000 persons.⁵

The initial reaction of many service delivery staff to CMS was one of skepticism and suspicion, as it came on the heels of a major workforce reduction. Simultaneously, a policy of devolving employment policy responsibilities from the federal to the provincial governments was being pursued. This added to the concerns of service delivery staff for their own job security. CMS, although being touted as an aid to service delivery by improving the effectiveness of program targeting, was viewed as a possible replacement for the case management approach. In the minds of some, it was viewed as an expert system that could replace counselors in HRCCs as a way to help the national headquarters achieve its goal of reducing full-time staff equivalents by 20 percent.

Despite the unfortunate context, focus group tests proceeded as planned. Interviews with HRCC staff after exposure to CMS features in focus groups indicated that most participants could imagine themselves working with a refined version of the software. Despite this majority view, there were pockets of resistance to the CMS approach that included two distinct camps: the "philosophically opposed" (or "Luddites," about 5 percent of participants) and the "threatened pessimists" (about 33 percent).⁶ The former group saw CMS as a challenge to their counseling methods, while the latter group feared CMS as a replacement for their services. Nonetheless, some constructive suggestions did surface from the focus group participants. These included the following:

- 1) Change the name Client Monitoring System, especially the word *monitoring*, which was viewed as threatening to both staff and clients because it implied "Big Brother."
- 2) Link CMS data with other key HRDC systems in various stages of development.
- 3) Ensure that management and service delivery staff have a shared understanding of how CMS would be used in improving the day-to-day operations at the HRCCs.

As a consequence of the focus group testing, an "alpha" version of the system was developed. The system name was changed from CMS to SOMS. Attempts were also made to link SOMS with other data and

accountability systems. However, these efforts failed because of incompatibility with the older data structures.

In 1996, after the successful business case presentation of SOMS to the Project Review Committee, Strategic Policy, and Systems branch formed a joint partnership to further SOMS development. Later that year, "beta" focus group tests of the SOMS system were planned for 10 of the 30 alpha HRCC sites. Around this time, devolution of employment programs from the federal government to the provinces was being done through bilateral agreements. As provincial management took over, several of the selected beta test sites dropped out of the plans for focus groups. In the end, only two of the selected local HRCC offices were left to participate in the beta tests, which went ahead in late 1996 and early 1997.

As a result of the sharp decline in the number of beta test sites, the methodological design for live system testing was modified. The test period was shortened to include only a comparison of the pretraining questionnaire data against that collected one month after system training.⁷ Focus group participants were positive about the quality of the information, the organization and presentation of client data in easily navigable screens, and the high level of security for confidential information. On the negative side, they downplayed the value of SOMS in helping to improve the quality of their work with clients. They also expressed concerns about the reliability and completeness of the data.

Despite the sometimes negative perceptions of SOMS's service delivery tool, in a March 1997 presentation of the system to an Assistant Deputy Minister with primary responsibility for all HRDC training and employment programs, the Assistant Deputy Minister suggested that SOMS replace the existing antiquated data-entry processing system used by local HRCCs. However, SOMS was not designed as a data-entry system. The time and resources needed to make the necessary changes were judged too large. In addition, resistance to a new system during a period of high uncertainty with respect to HRDC's role in the local labor market was likely to be strong. Rather than risk the entire project, which had in early 1997 received Treasury Board support and multiyear funding as an accountability system for HRDC, efforts were turned toward marketing and implementing SOMS's business application tool.

SOMS'S BUSINESS PLANNING TOOL

The other component of SOMS, its management reporting/accountability tool, was still relevant in a devolved department and development of this component, including the maintenance and updating of the relational database that supports the tool, continued. This component loaded directly on the end user's desktop computer and permitted managers and analysts to review summarized group data for the 10.8 million clients at various levels of detail and for different outcome measures. All of the national data were available at a glance in either spreadsheet or graphical format. Users could rapidly and easily explore the data of any multi-dimensional cube at any level of detail by filtering on the client (age, sex, education, unemployment compensation claim, etc.) and geographic dimensions. Users could also choose the outcome measure(s) to use in analyzing the effectiveness of service provided and its impact on clients served. The accountability portion of SOMS provided the manager or analyst with a powerful tool to review performance in order to make strategic decisions on where and to whom resources should be targeted.

Three data "cubes" (data sets) were developed and tested in the beta evaluation of SOMS: annual income information, employment insurance claims information, and intervention and results information. To build the cubes for analyzing grouped client data, data was first extracted and packaged in a format suitable for building the cubes by using software called Transformer. In Transformer, the analyst defines the data elements that need to be extracted from the source database and the important relationships between the elements. This forms the data model, which, after extensive testing for data consistency and correct relationships between the variables, is executed against the SOMS database to produce a number of PowerCubes. Each PowerCube contains a selection of extracted data, structured to show defined relationships, and stored in a proprietary format.

A six-week pilot test of the business application tool was conducted in Ontario during the summer of 1997. The test revealed that while the software was not as user-friendly as other "spreadsheet/analysis" software used, its graphical interface was far superior. Moreover, in comparison to other data sources available, SOMS was found superior,

as was the data quality and its organization. Shortcomings were noted in SOMS's geographic structure, the presence of "stale" data, and its querying ability.

Since the testing of SOMS's business application ended, agreements were reached with an HRDC partner—The Canadian Labour Force Development Board—and two provincial governments to test SOMS's business application tool on a trial basis. Other provincial governments also expressed an interest in testing SOMS.

SOMS'S PREDICTIVE MODELING COMPONENT⁸

The predictive modeling component, which slowed SOMS's acceptance by frontline delivery staff, was the system's Achilles' heel. The predictive models were designed to calculate which of the many programs and services delivered by HRDC had the best probability of improving the employment and earnings prospects for a client based on their sociodemographic characteristics and their past history of employment, earnings, and service receipt. The models were intended to allay criticisms often directed by local managers with respect to evaluation studies; namely, that although relevant to making policy decisions at a national level, such studies were viewed as irrelevant to frontline staff making day-to-day service delivery decisions.

To develop predictive "what works best" models at the level of the client, it was necessary to reorient the standard program evaluation strategy. In a traditional quasi-experimental evaluation, the net program effect is computed in a statistical model as the difference between the labor market outcome of program participants and nonparticipants, while adjusting for differences between the two groups. In the SOMS approach, all of the previous interventions received by a client are also included in the statistical model as independent variables, along with variables that measure standard sociodemographic variables, and various periods of elapsed time since the interventions were provided. Lagged dependent variables were also used as predictors of outcomes. Finally, to make the model relevant at the level of the individual, a number of interactive terms were also added using a "stepwise" regression procedure.

From each of the four regression outcome models specified—earnings, weeks of employment, savings in unemployment compensation, and probability of employment—predictive models were then developed for each of 25 interventions identified. To determine what works best, the predictive models use the regression outcome for a particular individual and increment the intervention by 1 (employment interventions) or by a specified typical number of weeks (training interventions) and estimate the outcome. The difference between the predicted and regression outcome measures equals the value of an additional unit of the intervention. By comparing the effect of each of the 25 interventions for any one individual, it is then possible to say which intervention will have the best effect.

While this approach was judged theoretically sound by leading econometricians, considerable difficulties were encountered in attempting to arrive at findings that could be generalized to the population and adequately differentiated between competing interventions. In early model rounds, although it appeared that the models could isolate the best intervention for a client, the predictive models often resulted in a majority of the clients (as much as 70 percent) being targeted to the same intervention. Furthermore, each of the outcome models tended to favor a different intervention.

Several refinements were adopted to improve the ability of the models to discriminate among alternative interventions. The revised models were able to identify more than one favorable intervention for each outcome, but confidence intervals for the program effect estimates were too large to precisely state which intervention was best. Moreover, the effect estimated for a number of the service outcomes was not statistically significant. While our efforts did not yield a tool to assist in service delivery, a number of findings which arose from the modeling efforts are important to consider.

First, in attempting to develop participation models to account for self-selection bias, it was found that there were such extreme differences between those who were past clients of HRDC and those who had never received service, that the participation models could not be built. The inability to construct a comparison group that both had not received an intervention at some time since 1987 and resembled those who did receive services strongly suggests the existence and operation of dual labor markets in Canada. That is, distinct markets for workers who are

usually job-attached and radically different markets for those who rely on the public employment service to assist them during jobless spells.

Secondly, missing data on key variables for large proportions of the sample populations resulted in large and significant bias in the estimation of program effects. This finding illustrates the importance of valid and complete client data entry in administrative systems, especially for those variables which are strong predictors of success. It also suggests that it would be useful to modify administrative data systems to capture certain information that, although not essential for program administration per se, is highly relevant in measuring success and maintaining accountability.

A number of reasons were postulated for our inability to predict reliably what would work best for a particular client in a given labor market area. In addition to the phenomena reported above, other potential reasons, which are backed up by the analysis conducted and/or the empirical research, suggest that the problem may result from the presence of unknown or unmeasurable attributes of clients, i.e., unexplained heterogeneity. In effect, people differ with respect to certain behaviors in ways that we cannot comprehend or model using available data. Also, individual programs have become more heterogeneous over time due to 1) dilution of selection criteria, 2) increasing devolution of service delivery from the federal to provincial governments, and 3) tailoring of interventions to match the characteristics of local labor markets. Increasing variation in the content and intensity of programs delivered can, by itself, result in imprecise estimates of intervention effects since the interventions themselves are imprecise.

Finally, the unavailability of precise cost data means that a cost-effectiveness ranking of net impacts for alternative interventions cannot be produced. If reliable cost data were available, the uncertainty about program referral resulting from overlapping confidence intervals might be greatly reduced.

RECENT SOMS DEVELOPMENTS

SOMS has moved far beyond the prototype stage. It is a fully tested, leading-edge, multifaceted accountability and targeting system

ready for wide-scale deployment throughout HRDC. Increased use of SOMS by national and regional headquarters for quickly constructing participant and/or comparison group samples affirmed it to be a reliable database for quick sample design and construction. Heightened interest in the SOMS business application tool by provinces and regions lent strength to the planned conversion of SOMS's programming code and an update of the SOMS database.

SOMS modeling revisions, which were completed in early 2000, succeeded in dramatically narrowing the confidence intervals, thus permitting much more precise statements about what works best. As a result, the SOMS modeling component was much more useful than at any prior stage of development. However, recent developments regarding the use of personal information for research and evaluation purposes that entailed the linking of databases from various sources have slowed the development pace, as multiple approvals are required by senior officials in more than one federal department or agency. Since patterns of program participants change over time, model estimates of what works best for whom have a finite useful lifetime. Unless the required approvals are sought and granted to build a new SOMS relational database system with refreshed current data, SOMS's potential as a service delivery and resource allocation tool will be lost.⁹

CONCLUSION

In the development of any accountability and targeting system, the highest importance must be placed on developing a reliable and credible database. If the results are to be meaningful and accepted, the data foundation must be trusted. Nonetheless, even after the best efforts to achieve this ideal, data anomalies will crop up in a system where the data is subdivided in so many ways to produce program effect estimates.

Building a system to meet many competing needs across a large organization is a challenging task. Constant testing and validation must be done to ensure that needs are met in terms of functionality, simplicity, and system compatibility.

Sometimes, as was found with SOMS, a system can be so technologically advanced that it is hard to link it with older systems. A new system can also pose a threat to the status quo and, as a result, be cast in a bad light or discredited entirely. Sufficient attention must be given to such factors for proper planning of a system with the size and complexity of SOMS. However, even the best planning cannot foresee all contingencies, and timing may play an overly significant role in deciding the fate and acceptance of a system.

Finally, “what gets measured, gets done (and gets attention).” In HRDC’s case, the important measures following the announcement of the new accountability system for the Employment Benefits and Support Measures (EBSM) were the short-term (3-month postprogram) unemployment insurance savings and return to work. SOMS reported on both in annual time increments. Instead of focusing on SOMS as a short-term EBSM outcome monitoring system, an effort was made to simply add that functionality to SOMS’s other features. In retrospect, concentration on a simple outcome monitoring system would probably have had the greatest effect on improving the acceptance of SOMS at the field level. However, besides being wasteful of resources, a second monitoring system would have increased confusion in HRCC’s trying to determine which system was best.

To avoid systems proliferation, efforts focused on linking and partnering the SOMS effort with other parts of the HRDC organization. This was seen as a means of reducing the total number of systems in use, while simultaneously improving their impact on the clients served and the results achieved.

Notes

This paper does not necessarily represent the views and opinions of the Government of Canada, nor those of the Department of Human Resources Development Canada.

1. In 1995, with the introduction of a revised program structure, Employment Benefits and Support Measures (EBSM), a formal accountability structure was introduced, requiring HRDC to report annually to Parliament on EBSM performance in meeting its short-term objectives of generating employment and saving employment insurance funds. Medium-term measures of employment stability, in-

come enhancement, etc. were also specified, but reporting on these measures needed the EBSM to have been in operation for a number of years before measures could be taken.

2. These points were brought out in various presentations made by the SOMS development team and were reinforced by the findings from focus group testing of various SOMS components over the 1995–1998 period.
3. A client is defined by SOMS to be anyone who had an employment insurance claim and/or a training or employment intervention at some point since 1987.
4. At the time of the focus group testing, and until early 1996, SOMS was called the Client Monitoring System (CMS). Focus group testing revealed the need for a change in the system's name.
5. In 1994, the Department of Employment and Immigration Canada (EIC) was reorganized as part of a major restructuring of the federal government. EIC lost its immigration component. All or part of four other federal departments were added to the remaining EIC. The newly formed HRDC accounts for almost all of the federal labor market and social programming. With spending of almost \$70 billion annually, HRDC accounts for one-half of total federal government spending.
6. Human Resources Development Canada (1995).
7. In each of the two offices, one manager was separately trained in using SOMS's PowerPlay business application.
8. Full model details are provided in the appendix.
9. In May 2000, the SOMS database was wiped out in response to concerns raised by the Office of the Privacy Commissioner regarding the extensive data holdings of HRDC.

Appendix

Details of the Modeling Approach for SOMS

In the past, program evaluations undertaken by HRDC have focused on determining the net effect of a program on a particular outcome indicator, by the use of a pre/post comparison group methodology and the estimation of a regression model, which took the form of

$$(A.1) \quad Y_i = \beta_0 + X_i\beta_1 + P_i\beta_2 + \mu_i$$

In this equation, the dependent variable Y is the outcome indicator; β_0 is the intercept term; the vector X_i contains the environmental and demographic variables of the program participants and comparison group members, and β_1 denotes their coefficients; P_i is a 1,0 variable indicating whether the individual participated in the program or not; β_2 is the marginal effect of a program; and μ_i is a random error term.

While the β_2 coefficient provides information on the incremental impact of the program being evaluated, it does not provide frontline HRCC staff with an answer to the question of whether the program would work for their clients, or, in the limit, for a particular client. Also, and as is normally the case, the delineation of individuals as participants or comparison group members is based on receipt or nonreceipt of a program. There may well be differences between the two groups in terms of the quantities of other programs received in the past. The implicit assumption of the standard equation for estimating program impact in a quasi-experimental research design is that the two groups are similar in terms of past programs and there is no bias in the estimate of the impact of the program under consideration.

In order to answer the question of which of the many available HRDC interventions would maximize the benefits received by a client, it was necessary to significantly alter the standard regression equation noted above. The heart of the SOMS predictive capability is a regression equation of the form:

$$(A.2) \quad UI_{i,95} = \beta_0 + UI_{i,(95-T)}\beta_1 + T_{ij}\beta_2 + X_i\beta_3 + I_i\beta_4 + Z_i\beta_5 + \psi_i\beta_6 + \mu_i$$

for $i = 1, \dots, 93,026$

where,

- $UI_{i,95}$ is unemployment insurance benefits paid in 1995 for the i^{th} individual.

- β_0 is the intercept term.
- $\mathbf{UI}_{i,(95-T)}$ is a vector of three values, representing lagged UI benefits paid in years $T - 1$, $T - 2$, and $T - 3$, where T is the year of the first recorded intervention on file for the i^{th} client. The coefficient vector for $\mathbf{UI}_{i,(95-T)}$ is denoted by β_1 .¹
- The vector \mathbf{T}_{ij} measures weeks since the last occurrence of intervention j for the i^{th} client. \mathbf{T}_{ij} consists of up to three elements, representing the distribution of times in separate linear pieces. This approach provides a flexible method of dealing with nonlinear relationships in the elapsed time since an intervention occurred in the past and the residual effect of the past intervention on the outcome indicator. The number of components and their precise definitions varies across interventions. The corresponding coefficient vector is β_2 .²
- The vector \mathbf{X}_i contains environmental and demographic variables, and β_3 denotes their coefficients.
- \mathbf{I}_i is the vector of intervention variables whose coefficients are β_4 . The vector, comprising 25 intervention variables (10 employment and 15 training), captures data on receipt of interventions over the period 1987–1994. Employment interventions are measured in terms of the frequency of occurrence over the time period, while training interventions are measured as the duration of training, in weeks. Both types are represented in the model by up to three component variables, where each component represents a piece of the distribution of the observed frequencies or durations as either a dummy variable or a linear approximation. The purpose of including components of this kind is to identify nonlinear relationships between the quantity of the intervention and the observed effect on the outcome indicator.³
- The variable denoted by Z_i captures the time elapsed between the receipt of the earliest intervention on record for the i^{th} client and January 1, 1994. The coefficient β_5 gives the relationship between this time variable and the outcome indicator.
- Terms representing the Kronecker product of the demographic, environmental, time, and lagged dependent variables (\mathbf{X}_i , Z_i , and $\mathbf{UI}_{i,(95-T)}$) with the intervention variables (\mathbf{T}_{ij}) are denoted by ψ_i . The coefficients of these interaction terms are denoted by β_6 .
- Finally, μ_i is a random error term.

The model is estimated by ordinary least squares (OLS). All variables except the interaction terms are forced into the model. For the interaction terms, a forward stepwise procedure is applied and only those interaction terms (components of ψ_i) which meet or exceed a 0.20 significance level are includ-

ed in the model. Before the stepwise procedure is applied, a total of 1,809 interactive terms are available to the model. Variables entered on a step may be dropped at a later point in the procedure if their calculated significance level falls below 0.22. The significance levels set for model inclusion and exclusion were chosen to achieve a balance between competing concerns. That is, to include a sufficient number of interaction terms to allow for differing estimates of what works best for clients, while at the same time trying to avoid the problem of multicollinearity. The resulting OLS estimates of the coefficients $\beta_1, \beta_2, \dots, \beta_6$ are denoted by b_1, b_2, \dots, b_6 .

The model described above can be used to estimate the reduction in UI benefits paid that results from the receipt of any given type of intervention. These savings can be assessed on an individual basis or on any level of aggregation (e.g., HRCC, region, province, etc.). The calculation requires several steps, as follows:

- 1) Estimate unemployment compensation receipt by the i^{th} person, $UI_{i,95}$ by substituting the OLS estimates b_1, b_2, \dots, b_6 for $\beta_1, \beta_2, \dots, \beta_6$ into Equation A.2 and evaluating the equation for the i^{th} person's characteristics ($T_{ij}, X_i, I_i, Z_i, \psi_i$).
- 2) Increment the value of the particular intervention j received by person i (T_{ij}). The intervention is increased by one unit if the intervention is measured as a frequency, or by the historically observed average number of weeks per occurrence if it is measured as a duration (e.g., the average duration of a training course).
- 3) Recalculate values of all explanatory variables ($T_{ij}, X_i, I_i, Z_i, \psi_i$) which depend on the value of the intervention.
- 4) Reestimate $UI_{i,95}$ using the recalculated explanatory variables and the original OLS parameter estimates (b_1, b_2, \dots, b_6).
- 5) The estimated effect of the intervention is then produced by subtracting the result of step 1 from that of step 4.

In addition to the savings in UI benefits paid, models were specified and tested for three other dependent variables: earnings in 1995, weeks of employment in 1995, and probability of employment in 1995. For the first two of these outcome indicators, a process similar to the one described above was followed to arrive at the final predictive equations. In the third case, a logistic regression model was used instead of OLS. The stepwise selection of interaction terms was different for each of the four outcome indicators.

Effects were estimated for 22 of the 25 interventions.⁴ The estimation, therefore, required 88 predictive equations—i.e., 4 outcomes by 22 interventions. Since the predictive models use interactive terms consisting of environmental and demographic variables specific to the client, SOMS can estimate

the impact of any one of the 22 interventions in terms of its predicted impact on a client's earnings, income, etc. In so doing, SOMS brings evaluative information down to the service delivery level and answers the question of "what works best" for a specific HRCC client.

The SOMS models are continually being refined and reestimated as a consequence of the dynamic nature of the data and the interventions upon which the SOMS models are based. New data, in addition to permitting reestimation of models, can also suggest changes in the formulation of the SOMS outcome models. Consequently, SOMS should be viewed as a dynamic model exercise which is sufficiently flexible to adapt to changes in the underlying data, as well as changes in HRDC's requirements for accountability and for information on what works best.

Appendix Notes

1. Previous SOMS models had only one lagged dependent variable term, defined for the period $T - 1$.
2. In previous SOMS models, T_{ij} was linear in construction, implying an assumption that the effect of past interventions was constant and not influenced by the time elapsed since receipt. The assumption of linearity runs counter to empirical literature, which suggests that the attenuation of effects is best depicted by a nonlinear curve.
3. Previous SOMS models accounted for possible nonlinear relationships between the intervention and its effect on the outcome indicator by using squared values of the main intervention variables, measured as either frequencies or durations.
4. Three of the interventions were residual categories for interventions that either were not captured specifically in the data (e.g., "other" purchased training) or occurred too infrequently to be modeled as separate interventions. These interventions were included in the models to compensate for their effects on the outcome indicators, but the process to estimate effects was not applied to them because such information would offer no guidance with respect to identifying an optimal intervention for a client or group of clients.

Reference

Human Resources Development Canada. 1995. "Focus Group Testing of the Client Monitoring System [SOMS] with CEC [HRCC] Staff." Ottawa: Human Resources Development Canada.

Comments on Chapter 10

Jeffrey Smith
University of Maryland

The Service and Outcome Measurement System (SOMS) represents an important advance in attempts to use statistical models to guide the assignment of participants in social programs to particular interventions. Such efforts are important given that what little evidence we have suggests that caseworkers may not do particularly well at this task (see Bell and Orr forthcoming; Plesca and Smith 2001; and Lechner and Smith 2001). The lessons that Human Resources Development Canada (HRDC) learned from developing the SOMS should provide useful guidance to similar efforts in the United States to develop the Frontline Decision Support System (FDSS) to guide the assignment of individuals to services provided under WIA (the Workforce Investment Act).

Using statistical models to target (or profile) participants into alternative services is the administrative innovation *de jour* in the public sector agencies that provide these services. Like earlier administrative innovations *de jour*, such as performance standards, statistical targeting is proceeding much faster in practice than the research base that should support and guide its development. One of the things that has remained foggy in much of the small literature on statistical treatment rules (STRs), of which SOMS and FDSS are examples, is the importance both conceptually and practically of the choice of variable on the basis of which to allocate individuals to services. This issue is discussed at length in Berger, Black, and Smith (2000).

In the U.S. unemployment insurance (UI) system, the variable used to target services is the predicted probability of UI benefit exhaustion. In the welfare-to-work programs described in Chapter 8, it is predicted levels of employment. The thing that makes SOMS relatively unique is

that it is explicitly designed to allocate participants to services based on the predicted impacts of those services rather than on expected outcome levels in the absence of service. By targeting based on the expected gains from alternative services rather than on predicted outcome levels, the SOMS maximizes the efficiency gains from the program, thereby providing (it is hoped) the largest bang per long-suffering (and in Canada they are indeed long-suffering) taxpayer dollar. These gains may come, of course, at some equity cost, as persons who would do poorly in the absence of expensive services will not receive those expensive services if their predicted benefit from them is small.

In addition to highlighting the conceptual value of basing an STR on predicted impacts, the SOMS also serves to illustrate the fact that constructing automated systems to assign participants to services represents a very difficult task indeed. SOMS and other similar systems represent an attempt to create an automated, ongoing, non-experimental program evaluation of a large number of alternative services. Automated in this context means that, once established, the system can reliably generate impact estimates without the frequent intervention of an econometrician. The parameters of interest in the evaluation implicit in the SOMS include predicted subgroup impacts for a nearly infinite number of subgroups defined by observable demographic characteristics, past service receipt, and past transfer payment receipt. The difficulty of the task is recognized when it is considered that we really do not yet have a robust methodology for conducting one-shot non-experimental evaluations (see, for example, the discussion in Heckman, LaLonde, and Smith 1999). Thus, it is not at all surprising that the SOMS developers had some troubles along the way.

In the remainder of my remarks, I would like to briefly discuss some specific issues that were raised in the course of developing SOMS. Some of these are mentioned in Chapter 10 and some are not. The first issue concerns the econometric method used to generate SOMS's impact estimates. This method consists more or less of "one grand regression." The implied comparison group is persons with low intensity services rather than persons who never receive any services but would be eligible for them. As Colpitts notes in the chapter, this was due in part to the fact that persons who were eligible but did not participate during the long period covered by the SOMS database were

an unusual group. This reflects the fact that the eligible population is more saturated with employment and training programs in Canada than it is in the United States.

This econometric strategy relies on the assumption of what Heckman and Robb (1985) call “selection on observables.” The idea here is that conditioning on observable characteristics—in this case quite a lot of them but with some notable omissions, such as years of schooling—will control for all selective differences between participants in any of the different services. This is a tall order for a non-experimental estimator in an evaluation examining only one service; it is perhaps an even taller one in the case of a system that seeks to generate credible impact estimates for more than a dozen services.

The virtues of this econometric strategy are threefold. First, it is readily understood by agency staff and can be explained in a simple manner (at least relative to other available evaluation strategies). Second, unlike the currently popular matching strategies examined in, e.g., Heckman et al. (1998) and Heckman, Ichimura, and Todd (1997), the SOMS econometric strategy uses off-the-shelf software. This is important for a system that, once launched, should require little in the way of expensive econometric maintenance. Third, it uses the data at hand, which do not include any obvious sources of exogenous variation that could be used as instruments in models that attempt to take account of selection on unobservables (e.g., motivation and ability) as well as selection on observables. It remains an important open question whether it would be possible in future versions of SOMS or in other systems of this type to adopt more ambitious econometric strategies.

The second issue worth raising is how to code the extremely heterogeneous interventions commonly offered to the disadvantaged and the unemployed. These include somewhat standardized services such as job clubs or job search assistance, as well as quite heterogeneous services such as classroom training in occupational skills. For the latter, should the system treat all classroom training of this type as one service, thereby leaving the case worker with substantial discretion about what specific occupation to have the participant train for? Or should it define training more narrowly, and attempt to produce separate impacts for specific occupational groups? This is complicated by the fact that not all types of training will be available in all localities. This issue

was not addressed in great detail in SOMS, where existing administrative categories were taken essentially as given. The optimal way to approach this question remains an issue for future research.

Related to this issue is the question of how to deal with multitreatment paths. In many programs of the type offered in Canada during the time that SOMS was created, participants may receive a sequence of services rather than just one. In some cases, these sequences may be preplanned, as when it is expected that job search assistance will follow classroom training for those who have not already located a job. In other cases, they may reflect a search for a good match between the participant and the service being provided. In these cases, the initial services received resemble the “tasters” built into the New Deal for Young People in the United Kingdom. These tasters explicitly allow New Deal participants to try out the different types of services offered by the program in the hope that additional information will lead to a better match between participant and service and thereby to a larger impact.

If the sequences are preplanned, a particular sequence of services (if sufficiently common) can simply be treated as a separate service, for which a separate impact estimate is generated. In the case where the sequences represent “search” among possible service matches, things become trickier, both in the predictive sense and in the sense of what services received by past participants to include in the impact estimation model. This aspect of the design of service allocation systems would also benefit from further analysis, both conceptual and empirical.

The final issue pertains to which set of services to attempt to estimate impacts for at all. Another way to think of this issue is how to incorporate prior information that certain services, such as orientation interviews or individual counseling sessions, are unlikely to have detectable impacts. Attempting to estimate impacts for services that are known in advance to have impacts too small to measure will reduce the credibility of the system and may lead to some embarrassing numbers (as indeed it did in some early versions of SOMS). It is important in designing these systems to focus on key services and not to attempt too much, especially in the first round of development.

In looking to the future it is useful to consider two lines of development for statistical treatment allocation systems such as SOMS in Canada and FDSS in the United States. The first is their transformation

into true expert systems. We already have a lot of knowledge from both social experiments and from high-quality non-experimental evaluations about what works and for whom. As noted by Manski (2001), the data sets from some of these evaluations could usefully be mined to extract even more information along these lines. This information is ignored in SOMS, which relies solely on its own internal impact estimates—estimates based on a methodology that emphasizes ease of automation over econometric appeal. Combining the evidence from other evaluations with the internal estimates from SOMS (or other similar systems) would substantially increase the likelihood that the system would actually fulfill its appointed task of helping to associate participants with the services that would benefit them the most.

Second, the service allocation component of SOMS or of other similar systems could be used to generate useful exogenous variation in service receipt that would then in turn reduce the bias associated with future internal impact estimates from the system. The basic idea is to introduce some randomization into the set of services recommended as a permanent feature of the system. The randomization would not require that anyone be denied service, only that some systematic variation in service receipt be introduced by varying the set of services recommended or their ordering in a way unrelated to the observable characteristics of the participant. In technical terms, the system would be creating an instrument that could be used to help in evaluating the program. Building this aspect into the system would relatively painlessly increase the credibility of, and reduce the bias associated with, the impact estimates used to guide service allocation.

In conclusion, it should be clear that statistical treatment allocation systems such as SOMS display great promise at improving the efficiency of service allocation in social programs. At the same time, the research base underlying these systems is woeful, a situation that the chapters in this volume only begin to address. Much remains to be done.

Disclaimer

The author worked as an occasional paid consultant to Abt Associates of Canada (now Applied Research Consultants) in the course of their work on the Service Outcomes

and Monitoring System (SOMS) under contract to Human Resources Development Canada (HRDC). The views expressed in this comment do not necessarily represent those of any person or organization other than the author.

References

- Bell, Stephen, and Larry Orr. Forthcoming. "Screening (and Creaming?) Applicants to Job Training Programs: The AFDC Homemaker-Home Health Aide Demonstrations." *Labour Economics*.
- Berger, Mark, Dan Black, and Jeffrey Smith. 2000. "Evaluating Profiling as a Means of Allocating Government Services." In *Econometric Evaluation of Labour Market Policies*, Michael Lechner and Friedhelm Pfeiffer, eds. Heidelberg: Physica, pp. 59–84.
- Heckman, James and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, James Heckman and Burton Singer, eds. Cambridge, England: Cambridge University Press for Econometric Society Monograph Series, pp. 156–246.
- Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64(4): 605–654.
- Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5): 1017–1098.
- Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics, Volume 3A*, Orley Ashenfelter and David Card, eds. Amsterdam: North-Holland, pp. 1865–2097.
- Lechner, Michael, and Jeffrey Smith. 2001. "What Is the Value Added by Caseworkers?" Unpublished manuscript, University of Maryland.
- Manski, Charles. 2001. "Designing Programs for Heterogeneous Populations: The Value of Covariate Information." *American Economic Review* 91(2): 103–106.
- Plesca, Miana, and Jeffrey Smith. 2001. "How Can We Improve Public Employment and Training Programs?" Unpublished manuscript, University of Maryland.

Targeting Employment Services

Randall W. Eberts
Christopher J. O'Leary
Stephen A. Wandner
Editors

2002

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Targeting employment services / Randall W. Eberts, Christopher
J. O'Leary, Stephen A. Wandner, editors.

p. cm.

Proceedings from a conference held on April 29 to May 1,
1999 in Kalamazoo, Michigan.

Includes bibliographical references and index.

ISBN 0-88099-244-1 (cloth : alk. paper) — ISBN 0-88099-243-3
(pbk. : alk. paper)

1. Unemployed—Services for—Government policy—United
States—Congresses. 2. Employment agencies—United States
—Management—Congresses. 3. Manpower policy—United States
—Congresses. I. Eberts, Randall W. II. O'Leary, Christopher
J. III. Wandner, Stephen A.

HD5875.T37 2002

362.85'84'0973—dc21

2002024572

© 2002

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by Jan Underhill.
Index prepared by Nancy Humphreys.
Printed in the United States of America.