

Making two plus two equal five

Adding value through the use of internal data in evaluation research

Anne Sharp

This paper examines ways in which internal data can be used in the evaluation research through data matching. It illustrates this through two case studies involving a three-year longitudinal and a cross-sectional evaluation of two large-scale Australian assistance programs. Discussion centres around ways in which internal data were used to assist the interview process, identify areas where 'claimed' (what the respondent says) and 'actual' (what the internal data show) data differ, and identify non-response issues.

The paper also examines how internal data can be used as an intermediate platform against which the results of two independent surveys can be reconciled. This is useful as many surveys do not use the same measurement scales or cover the same topics. The use of internal data can provide some commonality between the two, again adding value in the overall evaluation process and allowing the organisation to better utilise existing data.

Introduction

In the last decade, there have been dramatic improvements in the availability and integrity of information kept by organisations on their clients. This has resulted in a demand for more innovative practices in evaluation, as well as market research, to take advantage of this wealth of information. Increasingly, the demand is for practices that contextualise the data obtained in new evaluation work against already held internal data and past evaluations.

Australian governments are increasingly looking to maximise the value they get from evaluation research. There is a movement towards developing a few key streams of research, with the expectation that external providers will work together and that evaluations will dovetail and not be duplicated. This is evidenced by the increasing call for consortia to tender for evaluations, the appointment of a panel of providers rather than going to open tenders, and the move away from ad-hoc evaluation projects.

As funds are reduced, more internal government parties have a vested interest in a single piece of evaluation research, so that one study often has to cover several issues at once. Despite this, there has been only limited use of data matching aimed primarily at improving efficiencies in the research process and adding value to the research findings. This paper may help in this transition process, since it fits in with

Anne Sharp is a Senior Research Associate at the Marketing Science Centre, University of South Australia. She can be contacted at anne.sharp@MarketingScienceCentre.com. Her background is in market research, but for the last five years she has been involved in evaluation as an external consultant.

this more efficient way of conducting evaluation research.

Data matching involves merging selected items drawn from data from one source with data from another, using a common intermediary variable, typically a person's identity. In the context of this paper, we are referring to merging internal client data and histories with responses from external evaluation surveys. This is done through use of an intermediate variable, attached to both the respondent's internal file and survey answers. The identifying details (name, address, phone number, etc.) are then removed from both files and they are merged, using the unique identifier. This allows for data from the different sources to be compared, without compromising the anonymity and privacy of the respondent.

Government departments have been data matching for some time in countries including Australia. However, most of this data matching has been done to catch illegal behaviour such as social security or tax fraud. Privacy laws in many countries deter managers from considering such uses of data matching in evaluation (e.g. Hughes 1991)

Our experience is that, in some instances, the respondent will deny the categorisation given, perhaps because they cannot recall accurately or because they do not wish to fully disclose information.

and in other cases, managers' misuse of data matching has resulted in serious legal ramifications (Glen 1992). If used ethically and within the constraints and strict protocols of Privacy Act confidentiality provisions, the Market Research Society code of conduct and other acts or contractual obligations to protect the respondent and how data is used, data matching can lead to improvements in the quality of the data, as well as time and cost savings for both the organisation undertaking the evaluation and the respondent.

Another development in government departments is the setting up of evaluation processes that seek triangulation of findings – i.e. running multiple studies that look at the same issues from several perspectives to verify and enrich findings. Where multiple research programs are occurring, results of each can then be reconciled against the others through an internal database. This can be done, once again, via the unique data matching identifier, generated to protect the actual identity of the individual respondent. Such studies have been heard of before in social welfare settings in Europe (e.g. Torelli & Trivellato 1989) and in the United States (e.g. Radner, Kilss & Scheuren 1988), but are relatively new in Australia.

For the majority of organisations (including those in the public sector), the key problem remains that of having a wealth of data, but no information

about what it all means. While many organisations have taken advantage of new, cheaper ways of mass data collection to update their internal databases, most have failed to sit back, reconcile all the work they have done and understand how such information can be best used.

This paper gives examples of organisations in Australia that have taken the opportunity to use the data that exist more proactively and to reconcile them against existing knowledge. This paper is written from the perspective of a commercial evaluation research organisation, collaborating with an assistance provider that is evaluating a program. It is designed to give practical suggestions for other evaluation researchers.

The case studies

The first case involved a national, longitudinal evaluation. The evaluation required information to be obtained from families on issues such as financial management in the household, perceptions of role swapping and role sharing, and perceived barriers to achieving the respondent's desired level of workforce participation. Standard demographics were also obtained. The evaluation took place over three years (1998 to 2000), with the same 1300 respondents, where possible, being re-contacted in each of the research phases. The first phase had a significant qualitative component where respondents, and their partner, were interviewed in their home. This was followed by telephone interviews in subsequent phases. Opt-out letters were sent to all potential respondents and permission for access to their records obtained.

All but 300 of the respondents had a relationship with the organisation funding the research. For these people, internal information was available on the work status of the respondent and their partner, date and country of birth, number and age of dependent children, housing status, and their relationship with the organisation. These data were used in the interview process and then merged with the survey data in the analysis phase, after removal of respondent names and contact details, to increase the reliability and richness of the research findings. This required that steps be taken to keep survey data and information identifying respondents separate. This was done through the use of the aforementioned intermediate unique identifier number. The survey data and internal data could thus be matched through this number, once both sets of data had all personally identifying information removed. This allowed for a much richer dataset that still protected the identity and confidentiality of respondents.

The second case was cross-sectional. The fieldwork was conducted in 1999 and involved 25-minute telephone interviews with 1500 respondents nationally. The study sought to determine how

families chose between various service options offered by the organisation funding the research.

The likelihood of future choice behaviour was measured and the responses to various penalty and reward scenarios for the different options were elicited by self-report. This study required an overall measure of family decision-making and so both partners (if applicable) in the family were interviewed. Again, the evaluation had a significant in-depth, face-to-face interview component before conducting telephone interviews with families nationally. The same pre-evaluation opt-out process was followed.

Interviewing

There is well-documented research that shows respondents are poor at self-reporting their behaviour (e.g. Wind & Lerner 1979; Hu 1988). One of the aims of this paper is to illustrate how to get around the problem of relying on respondent recall alone. Conducting evaluations in a cost-effective manner can also mean conducting them in a research-efficient manner, while improving the quality of the data obtained.

Our work involved some unusual features such as the use of strict quotas to allow for sufficient post-survey sub-group analysis. These quotas are typically determined from internal client lists provided by the organisation sponsoring the evaluation. Our experience is that, in some instances, the respondent will deny the categorisation given, perhaps because they cannot recall accurately or because they do not wish to fully disclose information. You then end up with problems of meeting your quotas, as some respondents who are supposed to meet the membership criteria, claim not to. Since the organisational databases we worked with had recent records of clients' behaviour, it was possible to use a comparison of the internal and external data to assess the data integrity and the level of disagreement. Such feedback can be useful for the agency concerned, when determining the reliability of self-reported data and/or the accuracy of internal records. It also stops the problem of having to write multiple versions of the same basic report – one where the analyses are conducted according to claimed behaviour and status and one according to what the organisation's records show.

For example, in the cross-sectional survey, respondents were categorised according to their stated current relationship with the assisting organisation. It could be expected that these would be fairly high involvement relationships for the respondent, as they involved financial assistance. Because of this, we would expect little discrepancy between the respondent's claimed relationship (i.e. 'I am a customer of X services') and that shown by the organisation's records. However, when the claimed categorisations were reconciled back against internal data, almost 32% of cases were in dispute – what the client said and what the organisational records

showed did not agree.

Respondents were identified by the agency as being either a 'client' or 'not a client' of its services. This was done from internal records. In the survey we conducted, respondents were also directly asked, during the interview process, which agencies they used and their client status. Almost 50% of respondents who stated they were 'not clients' were actually on the client list of the organisation. The issue extends further. When respondents were classified into income groups according to their stated household income and then the same exercise was repeated using organisational data, there was a discrepancy in the classification in almost a third of cases.

Such an exercise can clearly be very useful for examining the reliability of survey and internal data. It does pose problems, however, in knowing whether to attribute the lack of agreement to internal record error, poor respondent recall, or lack of disclosure. We tackled this problem by holding a workshop with organisational staff where, after discussions, we attached probabilities to the different possible explanations for the discrepancy. We also used qualitative interviews to assess the extent of accurate client recall of behaviour and areas where they were more likely to not want to discuss issues openly in a telephone interview. This gave us some guidance as to which was the more accurate data source for each variable of interest.

The second unique aspect to such data matching is the reliance on respondents being able to recall what, to them, might be quite low-involvement information such as the specific names of services

Too much attention is placed on sample size and the associated margins of error, rather than examining the representativeness of the respondents.

they use. Once the study has been conducted, data from the organisation can be matched back to the respondent's survey data via the intermediate unique identifier. In such instances, we often find that the level of respondent recall error is high. Thus, if such data matching techniques are not used, it may result in many respondent cases being unable to be used, as respondents do not fit the criteria they claim. What is of more concern is that it may result in lost data in the instances where respondents misclassify themselves and so are excluded from being asked certain questions that they should have been asked.

An ability to account for non-response

All evaluation findings need to be clearly defensible. Recently there has been discussion of how non-response is often overlooked when considering the representativeness of research (Gendall 1999;

Fogliani 1999). Too much attention is placed on sample size and the associated margins of error, rather than examining the representativeness of the respondents (Fogliani 1999). In this area, the use of organisational data combined with survey data has a great deal to offer. In survey work where such data are not available, little can be found out about respondents who refuse to participate in the evaluation. The common practice of asking demographic questions at the conclusion of the interview means that virtually no data are available on respondents declining to be interviewed, or those dropping out part way through the interview. However, if the interview is done utilising organisational data on the individual, something can be known about non-respondents, even though they decline to be interviewed. Their privacy is still protected as the data used need not contain their contact details – it is simply a matter of creating a separate file with the internal data on respondents declining to be interviewed. The analyses can be done in-house, removing the need to have information passed on to any other party.

Information on non-response is a very desirable aspect of an evaluation when the results are used to inform future policy. The results can be objectively examined for non-response bias on any or all of the dimensions on which organisational data are held. From this, corresponding weights can be calculated

In terms of privacy, the evaluation practitioners have an obligation to ensure that confidentiality of respondent identity is maintained, and the client organisation has an obligation to maintain the privacy of the respondent's records.

if necessary to bring the findings into line with the overall target population for the evaluation.

In the longitudinal evaluation, the investigators were able to use the internal database to track and measure non-response caused by attrition. By drawing a sample from the parent database that mirrored that used to recruit the panel, and then tracing the same respondents through the three years, we could see at each stage of the evaluation if systematic non-response was occurring amongst certain respondents. For example, we found that there was greater non-response amongst those respondents who were renting. This result was expected due to renters' higher mobility and the ease with which we could lose contact with them. We also were able to identify non-response by certain ethnic groups, mainly due to the language demands of the interview process. This ability to have a profile of the database that matched that of the panel was critical in producing defensible non-response analyses.

Further analyses

Much evaluation reporting has been based on the respondent's claimed status. However, this is changing, largely because the concerns about privacy issues are being resolved and the integrity of

internal databases is improving. In terms of privacy, the evaluation practitioners have an obligation to ensure that confidentiality of respondent identity is maintained, and the client organisation has an obligation to maintain the privacy of the respondent's records. Such obligations are clearly outlined for members of evaluation and market research professional bodies (MRSA 1997). Increasingly, our evaluation work is using actual respondent status, as shown by internal records, as the basis for respondent classification prior to interviewing. For example, in our first evaluation study we used the type and number of the organisation's services used by the respondent as a basis to determine which questionnaire they would receive. In the second evaluation, we used family income, geographical location and service usage history to classify respondents.

The increasing use of computer-assisted telephone interviewing has meant that complex questionnaires are far easier to administer. Respondents can be skipped in and out of questions based on their responses or database information, without any reliance on judgements from the interviewer. This ability to determine the flow of a questionnaire from data uploaded from a database during the interview, allows for the respondent to be asked correct questions. For example, if respondents are receiving assistance relating to housing,

questions on accommodation and housing can be asked without requiring them to recall the particulars of the services they are receiving. This keeps the questionnaire simpler, both for the interviewer and respondent. It also

cuts the interview time in that the interviewer does not have to spend time trying to establish whether or not the respondent meets the required criteria.

The only potential problem with this approach is that it relies on the data provided by the evaluation sponsor. Even with accurate data, there will be instances where respondents deny meeting certain criteria, either through a very recent change in circumstances not yet recorded in the database (e.g. ceased work, became pregnant, completed a tertiary qualification), or incorrect recall. There appears to be no easy solution to this issue. One of the mechanisms we have built into questionnaires to track such eventualities is a blank page at the end of the interview where the interviewer can type in notes to the research manager about where such problems have occurred in an interview. Also, when a question is reliant on an internal data record for it to be asked, we provide a closed-ended option of 'respondent disagrees with the internal records' to cover this option. The respondent is skipped out of the section if this response is selected.

A further advantage of utilising organisational as well as survey data is the subsequent analyses that are possible. As with most company databases, there is a tendency to have a wealth of demographic data on customers as well as some measures of behaviour

in relation to the organisation. For example, we had data on the type of services people had been receiving over time, the number of times they had used the services, and a demographic profile of the family. What is often missing, but of great interest, is the explanation behind these changing behaviours. Using survey data, coupled with internal data, we were able to gain a picture of each respondent's activities and reconcile this against the changes in service status and type. For example, we could determine the proportions of clients who ceased to be eligible because they no longer met the eligibility criteria, had become employed in full-time work, or had separated from their partner and hence gone on to other services. Yet these were events that the respondents, in an interview setting, were not able to recall or associate with the service changes.

We were able to look at the attempted program interventions, such as a change in the level of service availability, and determine what longer-term outcome resulted from this activity. These sorts of analysis are not possible from internal data alone. Only shifts in the services used are shown, not the reason(s) for the shift. It is also not possible to gain such insights with survey data alone, given that respondent recall has been shown to be very poor; the services being evaluated are complex, and the differences between service 'brands' may be so minor that the respondent does not bother to distinguish. Thus, such work allows us to formulate explanations for the trends seen in internal data.

Internal data efficiency

The matching of internal and survey data can also assist the organisation in deciding what data should be kept on their clients. Consider a desired behaviour of the respondent, such as loyalty to an organisation – or conversely, movement away from its programs if it is an assistance organisation. We can model this behaviour of interest and see if the internal or survey data have higher correlations with it. Regression models can be built from the two types of data to better understand the respondents' behaviour and what the best potential predictive indicators are. This can help us decide which data we should be collecting internally and which we should not be bothering with. Variables that have very weak relationships with the key behaviours of interest are less important to collect and track than ones that correlate highly. This process of prioritisation and elimination lowers the cost of internal data collection and database management. It also makes it possible to determine the type of survey data to collect (those most highly correlated with behaviours of interest), which reduces time and costs of external research.

Internal data are also useful when attempting to compare the findings of several surveys that do not have a common demographic set platform. This occurs if they have been conducted at different times, or consist of different populations of respondents, or if there is no overlap in the

questions and scales used. By matching the internal data back against both surveys, we develop a common platform from which the two datasets can be reconciled. The internal data can help highlight where the two populations differ in terms of demographics, or in their relationship histories with the organisation. This is becoming increasingly important as organisations strive to maximise the value obtained from their evaluation funds and to streamline research efforts to avoid replication. Such reconciliation can also be a means whereby data 'missing' from one dataset can be estimated through looking at a comparison group.

Implications

This paper has outlined some of the initiatives that are being taken to bring greater value to evaluation work and to utilise more fully the wealth of internal data that is kept by organisations. We acknowledge that this process can often be slow to emerge. Some of the key barriers to adoption are: privacy issues, liaison with the programmers responsible for internal data management to gain the needed data within the project timeframe, and the evaluation practitioners adjusting to an investigation that is new in its approach and may involve activities outside the strict contract they are used to operating under. However, if the initiative is taken, it can result in better research, a more complete evaluation and a less taxing interview for the respondent.

References

- Fogliani, M. 1999, Low response rates and their effects on survey results, The Market Research Society of Australia 28th National Conference, The Market Research Society of Australia, Adelaide.
- Gendall, P. 1999, Responding to the problem of non-response, The Market Research Society of Australia 28th National Conference, Market Research Society of Australia, Adelaide.
- Glen, R. 1992, 'Secret databases', *Canadian Datasystems*, vol. 24, no. 12, pp. 50-54.
- Hu, M. & Bruning E. 1988, 'Using prior experience to explain survey versus diary recorded usage data', *Journal of the Market Research Society*, vol. 30, no. 1, pp. 59-72.
- Hughes, G. 1991, 'Federal regulation of databanks', *Australian Accountant*, vol. 61, no. 2, pp. 87-90.
- MRSA 1997, *Market Research Society of Australia Ltd 1997 National Directory & Yearbook*, ed., MRSA, Sydney.
- Radner, D. B., Kilss, B. & Scheuren, F. J. 1988, 'Commentary: Interagency data matching projects for research purposes; The 1973 CPS-IRS-SSA exact match study', *Social Security Bulletin*, vol. 51, no. 7, pp. 22-31, pp. 56-57.
- Torelli, N. & Trivellato, U. 1989, 'Youth unemployment duration from the Italian Labour Force Survey: Accuracy issues and modelling attempts', *European Economic Review*, vol. 33, no. 2, pp. 407-415.
- Wind, Y. & Lerner, D. 1979, 'On the measurement of purchase data: Surveys versus purchase diaries', *Journal of Marketing Research*, no. 16 (Feb), pp. 39-47.