

Using Secondary data for evaluating community programs – opportunities and challenges

Ilan Katz, Social Policy Research Centre, UNSW

Evaluators have always relied on secondary data to supplement primary data collection. However the use of secondary data has increased over the past few years, firstly because developments in technology and the increase in managerialism mean that more and more data is becoming available, and secondly because evaluators (especially in Australia) are increasingly required to produce robust ‘scientific’ (loosely translated this means ‘quantitative’) results with fewer and fewer resources being made available for primary data collection by researchers. Because secondary data is already produced for other purposes, it is a cheap source for evaluators, and offers the potential of a large and varied data set which can provide insights into many aspects of the program.

In this paper I will explore some of the possibilities which are opening up for evaluators, but also point out some of the pitfalls in the use of secondary data. I will conclude by offering some suggestions about how secondary data may be used in future evaluations. I will focus here on evaluations of initiatives aimed at children and families, and in particular area based or community initiatives, which are becoming an increasingly important component of early intervention, and which lend themselves to the use of secondary data.

It is in principle easy to define secondary data – it is all data which is not collected by the evaluators themselves or purely for purposes of the evaluation. However we shall see that there are a number of grey areas and the distinction between primary and secondary data is not always very clear.

There are three basic types of secondary data, two external and one internal to the program;

- Administrative data is produced by organisations and agencies outside the program for their own purposes, but which can be harnessed by the evaluation. Eg. crime rates or admissions to hospital.
- Research or survey data collected by governments through or other agencies census panel studies etc for their own purposes. Eg socio-demographics, social capital measures
- Management information. This is data produced by the program itself for accountability purposes in the form of progress reports, process data, financial data etc

In this paper I will confine myself to the first two points, but management information is an equally important data source for evaluators, and throws up equally interesting challenges and opportunities.

Administrative data

There have been huge advances in the use of administrative data for evaluating government policies and programs. Perhaps the most significant development in this area has been the attempt by policy makers to move from process based measures of outputs towards the use of outcomes frameworks to manage policies and programs. Most public authorities in Australia and the rest of the developed world are now avowedly ‘outcomes focused’ and there is increasing use by the public sector of systems such as the *Balanced Scorecard* (Kaplan & Norton, 1996) or *Results Based Accountability* (Friedman, 2005) which rely on the analysis of outcomes indicators. These systems are all based on the measurement of outcomes through administrative data. The UK is using an outcomes framework to measure the success of its reform of children’s services (called *Every Child Matters*) (www.everychildmatters.gov.uk). In Australia, the AIHW published *A Picture of Australia’s Children* (AIHW, 2005) which draws heavily on administrative data to provide a snapshot of the health and wellbeing of Australia’s children, with a view to establishing a baseline which can track the various dimensions over time.

At an international level administrative data is being used to attempt to measure outcomes (and by implication the effectiveness of policy) in many areas of child wellbeing – a good example of this is the attempt by UNICEF to create a ‘league table’ of child deaths by maltreatment (UNICEF, 2003). This report shows the enormous potential of secondary data for understanding and measuring outcomes, but it also demonstrates the pitfalls and challenges of using data in this way.

Technological advances in data collection and storage are beginning to allow researchers to make much more sophisticated use of administrative data. One of the more interesting developments is data linkage – which uses statistical techniques to link different data sets and allows, for example, health and educational data to be linked. This methodology will allow much more fine grained analysis of the service use trajectory of different types of people in different situations, and will answer such questions as *What is the relationship between levels of visits to GP before age 3 with school attainment aged 10?*

Fred Wulczyn and colleagues from the Chapin Hall Center in Chicago (Wulczyn, Barth, Yuan, Harden, & Landsverk, 2006) have used secondary data on child protection systems and children in OOHC to map out the trajectories for different ages and groups of children, showing, for example, that by far the highest proportion of children first come into OOHC before age 1 and that rates are higher for African American children at every age. Their analysis is further strengthened by relating the administrative data to knowledge of child development.

None of these developments relates directly to the use of administrative data for the evaluation of specific policies or programs, but there is great potential for this use. This is particularly true for programs aimed at whole populations or programs whose

theoretical basis is the *Ecological Model* (Bronfenbrenner, 1979). These programs rely on community level or universal interventions which are designed to improve outcomes for the whole population (or the population of a particular geographic area). They are thus measured by changes in the population rather than changes in individual children and families who have received specific services or interventions.

In principle there are a large number of administrative datasets which could be used for these purposes. Many of the outcomes which early intervention programs are designed to promote are either measured directly or there are reasonable proxy measures for them. In brief the advantages of using administrative data for programmes like this are:

- It is cheap. As the data is collected anyway the costs are very low. Sometimes the data has to be cleaned or manipulated to make it usable but even so this is by far the cheapest way of collecting data.
- It is comprehensive. Administrative data covers the whole population and is therefore not subject to the usual problems of sampling which characterise primary data collection.
- It is long term. Many of the data sets are collected continuously or regularly over long periods. This means that evaluators can establish baselines from before program starts, and can continue to analyse data indefinitely as long as the data is compatible.

This is equally true of survey data. The increasing sophistication of data produced by the census and other surveys, be they cross sectional such as the Child Care Survey, or longitudinal such as the LSAC (Longitudinal Study of Australia's Children) can be harnessed for evaluative purposes and used in a number of different ways.

In Australia, the Families First Program in NSW was the first to develop an outcomes framework, and now virtually every state and territory has developed or is developing their own framework. The Council of Australian Governments (COAG) is also working on a program to develop a national set of indicators.

Families First is a universal “whole of government” program, and it was decided early on that the indicators should apply to the whole population of NSW. Thus the outcomes are measured for the whole population of NSW, irrespective of their contact with the program, and, more importantly, irrespective of whether the program in their area actually provides services which address that particular outcome. The policy makers decided that because the program is intended to bring together existing services and policies for under 8's, and not only provide direct services, it made sense to apply indicators to the whole population. In fact the indicators are internally analysed by Families First area (which are coterminous with NSW health areas as they were when the program was established in 2003 – the health areas have subsequently changed). The difficulty with this approach is that it will be difficult to attribute any changes to the program, and it is difficult to measure the effectiveness of the program as a whole or of any of the interventions it funds.

Administrative data can also be used for the evaluation of community or neighbourhood initiatives, but this use poses different and probably more significant

challenges than for universal or national programs. These developments have been made possible by improvements in the technologies and processes for data collection. More and more datasets are available at the postcode or SLA level, and these data can be used to measure administrative (and sometimes community) outcomes at the community level. In addition, the use of geo-coding can considerably enhance the use of community data for research and evaluation. For example it is now possible to track outcomes for preschool children in relation to their proximity to child care, playgroups, parks or other facilities.

The advantages of secondary outcome data for community evaluations are similar to those described above for national or state level data. For example, in the same way as national data provides comparisons with other countries, and state level data provides comparisons between states and territories, local administrative data provides information about local outcomes which can be compared to national averages or with information from similar neighbourhoods (eg postcodes with similar SEIFA scores or similar demographic profiles).

In addition to its use for strictly evaluative purposes, local data (be it administrative or survey data) can be used by program implementers as a planning tool. This is especially true if data is geo-coded so that the planners can identify the existing services and relate them to the needs of the community. There is now a well established literature on how communities can use data as a tool for understanding needs and gaps in services in order to establish community priorities and plan interventions. Two Australian examples of this are provided by *Communities That Care* (Williams, Toumbourou, McDonald, Jones, & Moore, 2005) – a community program aimed at preventing crime and anti social behaviour amongst adolescents, and the *Australian Early Development Index* (http://rch.org.au/australianedi/index.cfm?doc_id=6210) – an initiative which provides a post code level measure of school readiness of children in communities.

However there are considerable conceptual as well as practical challenges in using secondary data in this way. The most important of these is that there are very few datasets, of either administrative or survey data, which are available consistently across Australia at the postcode level or below. The census is really the only reliable data source which covers every postcode in the country. This situation contrasts with that in the UK, for example, where there are a raft of neighbourhood statistics available online for every local authority area, ward and even Super Output Area (equivalent to Statistical Local Area). These statistics provide data about the population as well as services, economic activity, crime, work, deprivation the physical environment etc, all available to researchers and the general public on the Office for National Statistics website www.ons.gov.uk. In Australia local data, where it exists, has to be bought from the ABS at great expense to researchers

But even if there were nationally consistent datasets available for all postcodes, there would still be a considerable challenge to use them effectively for evaluative purposes. One of the major problems is that the “community” which is subject to the intervention is very difficult to define. Mostly administrators use post codes, but these are only approximations of communities. In Australia some post codes are geographically huge – covering several outback communities which can be very far

physically, demographically and socially from each other. Even in urban areas there are many post codes which have mixed populations. Consequently measuring changes in the mean of any particular outcome has little value. Going below the post code level is preferable, but there are very few data sets available for SLAs. In addition, analysing data at too small a geographic level is problematic – very small changes can result in large changes of the mean. For example if one or two new families with large numbers of children are reported to Community Services, this can make a significant impact on the reporting rates for an SLA. Another consideration is that measures can change at the level of the community even if programs have had little effect on individuals. If there is considerable churn in the population, then changes in such factors as feelings of trust in the community are likely to reflect the socio-demographics of the new population rather than the effects of specific interventions.

However the definition of community and reliability of postcode data are not the only challenges to the use of administrative data for evaluative purposes; the indicators themselves have to be adequate. Moore and Brown (2006) cite the following criteria for using indicators:

Readily understood	State-Specific data are crucial
Timely	Grounded in science- and in social values
Available over time	Amenable to change
Information on subgroups is needed	Measured across domains
Positive outcomes need to be measured	accessible

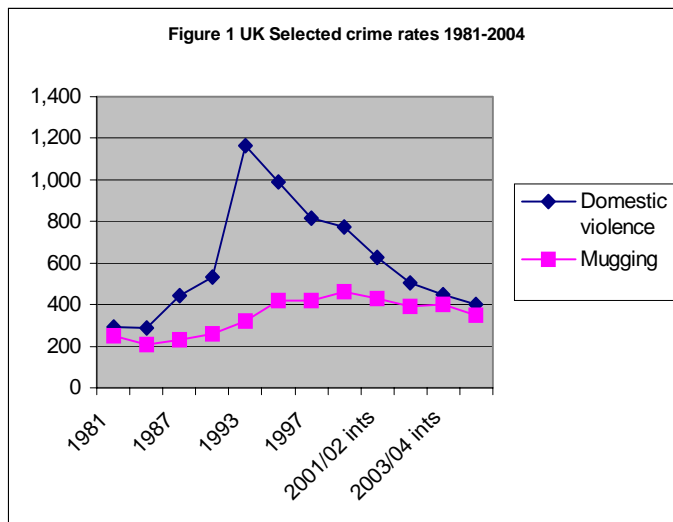
I will not discuss all these criteria, but many of those I mention here resonate with Moore and Brown. Our experience at SPRC over a number of programmes has led to the following list of criteria for the use of secondary data:

Consistency

This is a significant issue for many datasets because it means that the data definitions have to remain consistent over time and across the country. Because the primary purpose of administrative data collection is to facilitate current understanding of the population, data definitions often change. For example the definition of ‘crime’ and therefore crime rates can change if new offences are legislated for or if policy priorities change. As technology and systems develop, policy makers often want more fine-grained data in order to tailor policies to local (or national) needs, and this sometimes requires changing data definitions. For example data on ethnicity tends to change every few years as the population changes. Also new areas come into policy prominence – five years ago child obesity was not an issue in Australia but is now a very high policy priority. There is therefore often a tension between maintaining consistency so that changes can be accurately measured, and adapting indicators to measure the factors which are most important for policy makers. From an evaluation point of view this applies to factors relating to the program – at the beginning of programs it is important to measure implementation but towards the end factors relating to sustainability are more salient.

Another issue of consistency applies when the **meaning or significance** of data changes, even if the definition stays the same. This is particularly significant for data

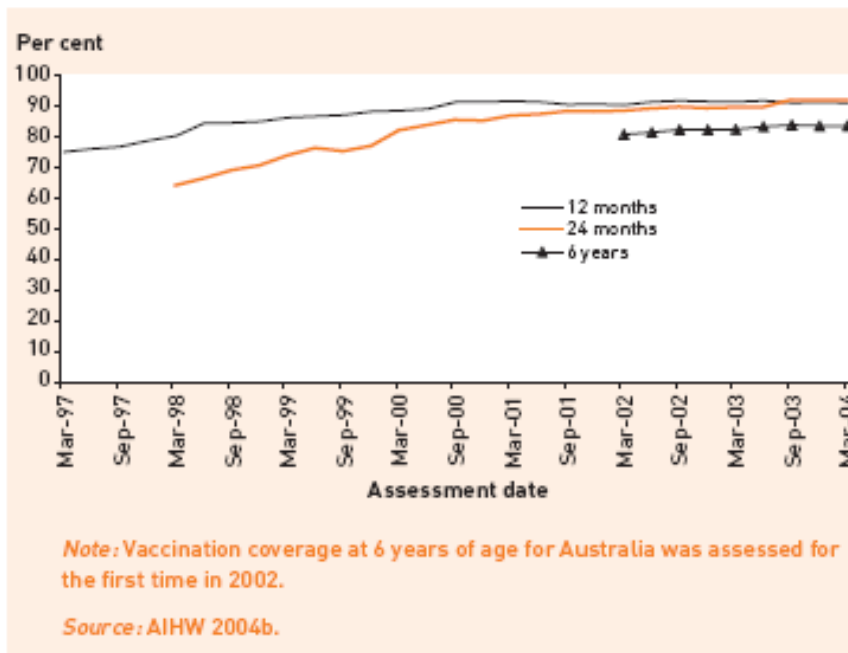
which is used as a proxy measure for outcomes. Meanings may change because of contextual issues or of policy changes. This is a particular issue when reported rates (of crime, hospitalisation, truancy etc) are used as proxies for prevalence in the community. For example hospital admission rates are often used as proxies for serious health problems, but these are very dependent on factors such as the availability of primary health care and the admissions policies of hospitals themselves. Whilst some of these factors may be 'smoothed out' at the national or state level, local variations in policy and practice can have a significant impact on these measures at the community level. Some of these issues apply even for panel data, which is a much more direct measure of prevalence. For example Figure I gives the figures for the UK rates for domestic violence and mugging. These are based on the British Crime Survey, which is a random survey of the general population (ie not based on reported rates to the police) it shows that domestic violence rates rose by 300% between 1987 and 1995 and that by 2005 was back to 1987 levels. The explanation for these figures is that during the late 1980s there was a lot of publicity about domestic violence, which had previously been a 'hidden' topic. Police set up domestic violence units and many NGOs began to publicly campaign for an end to DV.



Sensitivity to changes over time

Although, as indicated above, volatility in indicators is a problem, too much stability equally reduces the usefulness of the data. Some outcomes (or proxies for outcomes) are remarkably stable over time, and although they are of high policy relevance, are not useful as indicators because they do not provide data which is useful to policy makers. There are several reasons why indicators remain stable. One issue is that there is sometimes a *ceiling effect* where very high proportions of the population demonstrate a particular outcome. An example of this is vaccination rates. During the 1990s these would have been reasonable outcome indicators, but since 2000 or so they have remained more or less constant, and have probably reached their ceiling.

Figure 7.1: Trends in vaccination coverage, 1997–2003



Timeliness

One of the key practical challenges for evaluators is the need to align the data collection timescales with the logic of the program itself. Ideally the data will be collected at baseline, during the program, at the end of the program and then at regular intervals. That is fine for administrative data collected annually (such as admissions to hospitals). Even then it often takes months or even years for data to get into the public domain, and so secondary data sets usually have a ‘lag’ factor. In some cases the data collection is simply too out of sync with the program to be much use, even if the actual data set is directly relevant. For example the Stronger Families and Communities Strategy (SFCS stage 2) is a program which is being funded by the Federal Government between 2004 and 2008. Many of the outcomes developed by the federal government would best be measured by the census. [give examples] However the census is collected in 2001, 2006 and 2011. This means that it is virtually useless for the program – the 2006 census applies to a time when much of the initiatives funded by the Strategy are not fully functional, and since the evaluation ends in 2008, the 2011 census is also not helpful.

Degree of proximity (or ‘proxiness’)

Administrative data very seldom measures outcomes themselves. Almost invariably it measures a proxy of the outcome of interest. I have already discussed the issue of how contextual factors can influence the degree of proximity (ie the degree to which reporting rates reflect actual prevalence). Another issue is how closely **in principle** the measure is to the underlying construct. Domestic violence, discussed above, is very close, but much administrative data is far more removed from the outcome of interest, and are therefore difficult to interpret. Perhaps the most notorious of these is child maltreatment notifications. Not only are these notoriously dependent on local practice, recording processes, media attention etc, but it is also very difficult to interpret the meaning of changes in notification rates. Is an increase in notifications

good because it shows a higher level of awareness of child abuse, and a greater willingness to report, or is it a sign of program failure and an indication of higher rates of maltreatment in the community?

One problem with indicators which are not accurate proxies of the desired outcome is the creation of perverse incentives. For example it is known that dental health is a good indicator of children's overall health status. It is also easier to measure and applies to all children. It was therefore used as an indicator in the Children's Fund (a program aimed at improving outcomes for children aged 5-13) in the UK. But many of the programs created services to address dental health issues not because they were the highest level of local need, but because the program leaders wanted to ensure that their program successfully tackled the factor being measured.

Commonness

A key challenge relating to data sources at the community level is that the outcome in question must be reasonably common. Although mortality is a very good proxy measure of child health, and child death due to maltreatment is the best proxy for child maltreatment at the national level, these factors are useless measures for community interventions. Useful community measures must apply to a fairly high proportion of the population or they will have little value. This is a real challenge for evaluators who need to measure outcomes for sub-sections of the population such as Indigenous, disabled or CALD children. Even if data is available on sub groups, it is unlikely to be meaningful because there are too few in each community to incorporate into the analysis.

Accuracy vs simplicity

There is always a tension between simplicity and accuracy in the use of outcome indicators. As Moore and Brown (2006) note, policy makers and the public need indicators that are readily understood and simple, at least in principle. It is easy to understand that the rate of breast-feeding went up or down in a particular area over a specific period, even if the actual operational definition is more complex (eg proportion of children who are exclusively breast-fed by age 4 months and partially breast fed by 6 months).

However many comprehensive community interventions address a large number of outcomes in different domains (health, education, wellbeing, social capital, crime etc). The evaluator has three alternatives 1) to develop a large number of indicators to measure the various components of these domains; 2) to choose a smaller number of 'key indicators' to represent each domain or 3) to develop aggregate indicators which cover various sub-domains. All of these options have to balance simplicity with accuracy, and none is an optimal solution. A good example of this dilemma (although there is very little secondary data in this area) is *Social Capital*, which is commonly believed to consist of the following dimensions: *Trust; Community Engagement; Social Networks; Reciprocity* and *Shared Norms*. (Black & Hughes, 2001). It turns out that these sub dimensions are not all that closely related to each other – communities can display very high levels of engagement but low levels of trust. Thus

a composite measure of social capital is not very meaningful, and quantifying changes in such a measure is likely to show very little. On the other hand, tracking changes in all these dimensions is equally problematic and difficult to interpret, as is selecting one of them to represent all the others. In this case the best option would probably to go with multiple measures, but a judgement needs to be made in each case.

Sensitive to child development

Evaluators of programs aimed at improving outcomes for children have to factor in an additional level of complexity – child development. Risk (and protective) factors change as children develop, and so measuring changes over time requires an adaptation of the measurement to the children’s age. This is particularly challenging for early intervention programs which are aimed, at least in part, to prevent poor outcomes (eg anti social behaviour) or promote protective factors (eg active citizenship) in the future. Typically the logic model of such programs sets out short term outcomes (improved parent child relationships, school readiness etc) which are expected to lead to medium term changes (improved school attainment at primary school, less disruptive behaviour, improved relationship with peers) which in turn will lead to the desired adolescent or adult outcomes. But because the programs are comprehensive, it is also expected that they will lead to short term improvements to the whole community. So the question arises whether it is legitimate to measure anti social behaviour now (on the grounds that the whole community should benefit from these interventions), or should the evaluation confine itself to immediate outcomes of the primary target group – breast feeding, height percentiles, school readiness etc? Again there is no easy answer to this question. If the former approach is taken, then the issues of population churn, definition change and contextual transformation will undermine any chances of finding changes several years after the program ends, even if it has had an impact on its primary target group.

The meaning of averages

Most indicators are measured by tracking changes in the mean or median score for particular outcomes over time. However relying on averages can be very deceptive. Most comprehensive community programs aim to reach the most vulnerable and disadvantaged children in the community. But many programs become ‘colonised’ by middle class families who have the human and social capital to benefit from them. It may then be found that the community has improved in school readiness, but in fact what has happened is that some sectors of the population have benefited, and that those are not the groups targeted by the intervention. This is precisely what was found in the Sure Start impact evaluation (NESS (National Evaluation of Sure Start), 2005), although the **indicators** did show some improvements at the community level. One solution to this is to report on standard deviations rather than mean scores. In essence this is measuring equality of outcome in an area, rather than average outcome. Although this is closer to the expressed goals of most programs, it is a far more complex construct than mean scores, and more difficult to interpret changes in standard deviations over time.

Conclusions

This paper has tried to show the considerable benefits, but also the conceptual and practical challenges, of using secondary data for evaluations. It has focused on the use of administrative data in the evaluation of comprehensive community interventions with children and families. The use of administrative data in this context is still in its embryonic stage, and we still do not understand all the forces at play. Contextual issues such as public awareness, reporting rates, administrative procedures and collection processes all affect secondary data collection, and there are difficult choices to be made about what data sets should be used and how the data should be presented in order to maximise their benefit to the evaluation. Nevertheless the rewards are potentially significant. Administrative data, survey findings and management information all provide cost effective sources of data which would be prohibitively expensive and burdensome to collect via primary research methods.

References

- AIHW. (2005). *A picture of Australia's children*. Canberra: Australian Institute of Health and Welfare.
- Black, A., & Hughes, P. (2001). *The identification and analysis of indicators of community strength and outcomes, Occasional Paper Number 3*. Canberra: Department of Family and Community Services.
- Bronfenbrenner, U. (1979). *The ecology of human development : experiments by nature and design*. Cambridge, Mass.: Harvard University Press.
- Friedman, M. (2005). *Trying Hard is not Good Enough: How to Produce Measurable Improvements for Customers and Communities*. Victoria, B.C.: Trafford,.
- Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 75-85.
- Moore, K. A., & Brown, B. (2006). Preparing Indicators for Policymakers and Advocates. In A. Ben-Arie & R. Goerge (Eds.), *Indicators of Children's Well-Being: Understanding their role, usage and influence* (pp. 93-104). Dordrecht: Springer.
- NESS (National Evaluation of Sure Start). (2005). *Variation in Sure Start Local Programmes' Effectiveness: Early Preliminary Findings. Report of the NESS Programme Variability Study* (No. NESS/2005/FR/014). London: Institute for the Study of Children, Families and Social Issues, Birkbeck, University of London.
- UNICEF. (2003). *A league table of child maltreatment deaths in rich nations', Innocenti Report Card No.5, September 2003*. UNICEF. Florence: Innocenti Research Centre.
- Williams, J., Toumbourou, J. W., McDonald, M., Jones, S., & Moore, T. (2005). A sea change on the island continent: frameworks for risk assessment, prevention and intervention in child health in Australia. *Children & Society*, 19(2), 91-104.
- Wulczyn, F., Barth, R. P., Yuan, Y.-Y. T., Harden, B. J., & Landsverk, J. (2006). *Beyond common sense; child welfare, child well-being, and the evidence for policy reform* New Brunswick: Aldine Transaction.