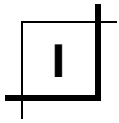

Secondary data analysis: an introduction

> All data are the consequence of one person asking questions of someone else. (Jacob 1984: 43)

This chapter introduces the field of secondary data analysis. It begins by considering what it is that we mean by secondary data analysis, before describing the type of data that might lend itself to secondary analysis and the ways in which the approach has developed as a research tool in social and educational research. The second part of the chapter considers the use of secondary data analysis in contemporary social research and introduces the results of a review of recent research output in the field.

Defining secondary data analysis

Numerous definitions of secondary data analysis appear in the literature, many with subtle differences which together suggest a lack of consensus about what is meant by the term. For example, one relatively straightforward definition of the secondary analysis of survey data was suggested by Hyman (1972: 1), as 'the extraction of knowledge on topics other than those which were the focus of the original survey'.

Other definitions of secondary analysis have emphasised its usefulness for exploring new research questions: 'the study of specific problems through analysis of existing data which were originally collected for another purpose' (Glaser 1963: 11); or: 'the further analysis of an existing dataset with the aim of addressing a research question distinct from that for which the dataset was originally collected and generating novel interpretations and conclusions' (Hewson 2006: 274). However, such definitions appear to disregard the potential of secondary analysis in re-analysing existing datasets with novel statistical or theoretical approaches

in such a way that: 'secondary analysis is the re-analysis of data for the purpose of answering the original research questions with better statistical techniques, or answering new research questions with old data' (Glass 1976: 3). One apparent area of consensus among those looking for a definition of secondary analysis is that it should involve the analysis of someone else's data: 'a collection of data obtained by another researcher which is available for re-analysis' (Sobal 1981: 149). However, this has been disputed as: 'even re-analysis of one's own data is secondary data analysis if it has a new purpose or is in response to a methodological critique' (Schutt 2007: 4127). Other researchers point to the ubiquitous credentials of secondary analysis as 'neither a specific regime of analytic procedures nor a statistical technique, [but] ... a set of research endeavours that use existing materials' (Kiecolt and Nathan 1985: 10). And emphasise its difference from primary analysis: 'which involves both data collection and analysis, while secondary analysis requires the application of creative analytical techniques to data that have been amassed by others' (Kiecolt and Nathan 1985: 10).

Given the rather subtle differences in the definition and interpretation of secondary analysis that we see here, it seems likely that neat distinctions between primary and secondary data will not always be possible (Dale et al. 1988). Such lack of consensus might leave one wishing to adopt a very general definition of secondary analysis such as that offered by Jary and Jary (2000): as 'any inquiry based on the re-analysis of previously analysed research data' (p. 540) or one such as Hakim's:

> [S]econdary data analysis is any further analysis of an existing dataset which presents interpretations, conclusions or knowledge additional to, or different from, those produced in the first report on the inquiry as a whole and its main results. (Hakim 1982a: 1)

Whichever definition one favours, secondary analysis should be 'an empirical exercise carried out on data that has already been gathered or compiled in some way' (Dale et al. 1988: 3). This may involve using the original, or novel, research questions, statistical approaches and theoretical frameworks; and may be undertaken by the original researcher or by someone new.

What are secondary data?

Secondary data can embrace a whole spectrum of empirical forms; they can include data generated through systematic reviews, through documentary analysis as well as the results from large-scale datasets such as the National Census or international surveys such as the Programme for International Student Assessment (PISA). Secondary data can be

numeric or non-numeric. Non-numeric, or qualitative secondary data, can include data retrieved second hand from interviews, ethnographic accounts, documents, photographs or conversations. In the UK, an excellent source of archived qualitative data with huge potential for secondary analysis is available through the Economic and Social Data Services (ESDS) Qualidata facility based at the University of Essex. Data available through Qualidata includes in-depth and semi-structured interviews, field notes and observations, as well as personal documents. The service provides support and training, as well as access to contemporary and classic studies of British society, such as the research papers and data for Dennis Marsden and Brian Jackson's 1962 study, 'Education and the Working Class' (see Appendix 1 for details on accessing the ESDS facility). In this book, however, our concern is with numeric secondary data only. (But see Hammersley (1997), Heaton, (1998) and Fielding and Fielding (2000) for further discussion on the methodological and substantive implications of the secondary analysis of non-numeric data.)

The potential for the secondary analysis of numeric data is huge. From a nation's population census to snapshot public opinion polls about the outcome of televised talent show competitions: 'Nearly every important area of activity and attitude in the British population has now been the focus of a major national survey' (Thomas 1996: 3). The range of numeric empirical data that are suited to secondary analysis would include:

- population census
- government surveys
- other large-scale surveys
- cohort and other longitudinal studies
- other regular or continuous surveys
- administrative records.

The next section will introduce these different types of secondary data, although a fuller discussion of specific datasets and data archives is provided in Appendix 1.

Population census

For example, the National Decennial Census or the School Census in the UK.

The National Decennial Census is arguably the gold standard of survey design. The resources needed to develop and administer the process dwarf those of other surveys. For example, the cost of the 2001 UK National Census was around £207 million in a cycle that lasted almost a decade (Office for National Statistics 2005). In 2003 the United States Census Bureau had a budget of $3.9billion and employed more than

860,000 workers in its ongoing data collection efforts (Sales et al. 2006). In England and Wales, an annual school census was introduced in 1997 which gathered data on the aggregate characteristics of young people in school. In 2002, this was extended to an annual *pupil*-level census, which itself became termly (three times a year) from 2007.

Government surveys

For example, the Labour Force Study or the General Household Survey in the UK or the General Social Survey in the USA.

Given the high level of funding and expertise that goes into the development of government surveys, they can often represent the highest quality data that are available. Their scope is generally large and their population can be highly representative, which aids robust inferences (Sales et al. 2006). They are likely to involve skilled teams of survey developers and statisticians and be administered by trained interviewers, so reducing the potential for interviewer bias (Dale et al. 1988). Often these surveys are produced at regular time periods and so can be combined to produce a type of 'synthetic cohort study' (Arber 2001: 276). For example, subsequent waves of the General Household Survey can be used to track a particular birth cohort through time. Although this survey does not track the same individuals, it does track the same cohort and so comparisons between different demographic groups may be made over a 5- or 10-year period, for example. One of the downsides of government-sponsored surveys is that the questions asked may reflect narrow contemporary policy interests, rather than topics that may be of direct interest to the social science researcher. Indeed, the concerns of many social scientists about the use and potential abuse of government-sponsored or 'official' data is well documented, and will be discussed in the next chapter.

Other large-scale surveys

For example, the British Social Attitudes Survey or the Programme for International Student Assessment.

Not all large-scale surveys are produced by the state. However, just like government-sponsored surveys, other large-scale surveys are likely to involve collaborations between professional statisticians and survey designers and are arguably of the best quality available. Their scale can also ensure that they are nationally representative and so enable rigorous comparisons between different social groups, potentially over extended periods of time. The level of resources behind some large international surveys can be quite staggering. For example, the institutes and teams behind the Programme for International Student Assessment (PISA)

study include a secretariat, which is responsible for the day-to-day management of the programme, a governing body, international contractors for each of the more than 50 countries involved in the programme, national project managers as well as subject matter and contextual question expert groups. PISA, which is funded by the OECD member countries, accounts for around 30% of OECD's education budget.

Cohort and other longitudinal studies

For example, the British Household Panel Survey and the Millennium Cohort Study in the UK and the High School and Beyond study in the USA.

Longitudinal data enable the researcher to look at continuity and change in behaviour over time, rather than just focusing on brief cross-sectional snapshots of an individual's life (Brooks-Gunn et al. 1991). Panel and cohort studies are forms of longitudinal analysis which are conducted by collecting data at a number of points in time from the same group of people. They share many of the positive features of government and other large-scale surveys but with the additional benefit of enabling researchers to monitor social phenomena over extended periods of time. In the UK, the Centre for Longitudinal Studies, based at London's Institute of Education, is responsible for administering three national cohort studies: the 1958 National Child Development Study (NCDS), the 1970 British Cohort Study (BCS) and the 2000 Millennium Cohort Study (MCS). Datasets from these studies, as well as support and training, are available for secondary analysis from the Economic and Social Data service (ESDS) (see Appendix 1 for more details).

Other regular or continuous surveys

For example, Gallup opinion polls, smaller scale academic studies.

These surveys may be smaller in scale than government-sponsored studies and are likely to have been undertaken perhaps by a small team of academic researchers, by or on behalf of advocacy groups or as a piece of market research or public opinion poll. Their quality may vary, as indeed might their potential for bias and generalisability. For example, they may represent non-random population samples, with the data collection undertaken by novice or untrained researchers or, in the case of some academic studies, by students (Hakim 1982a). Alternatively, commissioned public opinion polls may be of high quality and adopt the latest sampling techniques but may be more susceptible to question bias and, because of their very fast turnaround, poor question piloting. However, results from these ad hoc surveys can also be fruitful sources of data for the secondary analyst. Data of this type are best located from

social science data archives or opinion polls archives such as those of the Roper Centre in the USA. In the UK, it is a funding requirement that studies supported by the Economic and Social Research Council (ESRC) make their data available for archiving.

Administrative records

For example, prison or probation records, data on admissions to higher education.

This category includes data that have been collected as part of routine data management and administration activities, for example, the characteristics of individuals who apply for higher education courses or the number of students who achieve particular examination grades. Often these data are available in aggregate form and require limited reanalysis. However, these data may also take the form of administrative records such as hospital medical records, police and judicial records and so on. With these types of data, issues of confidentiality and access are likely to be paramount, as is the amount of work required to retrieve, prepare and analyse the data. Additionally, these data begin to blur the arbitrary boundary between numeric and non-numeric secondary data. Data retrieved from administrative records can be very powerful, both in the way they can be used and reported by the media, advocates or governments, but also in their potential for understanding social phenomena.

We will revisit each of these categories of secondary data later in the book; but, first, we consider the development of secondary analysis as a research tool and the extent to which it is currently used in contemporary social science research.

Development of secondary analysis

Secondary analysis has a long pedigree. In 1790 the first national population census was undertaken in the USA, followed in Great Britain in 1801. The potential of these data for secondary analysis and their contribution to the social sciences is exemplified by Booth's work on occupation patterns that were derived from secondary analysis of the 1801–1881 UK Censuses (Booth 1886). By the end of the nineteenth century, the large-scale studies of urban poverty that were pioneered by Joseph Rowntree in York and Charles Booth in London marked the start of the social survey movement and the wealth of opportunities it afforded for secondary analysis. Official records were also put to use for secondary analysis, most notably in Durkheim's important research into the sociology of suicide (Durkheim 1952) which is considered in more detail in the next chapter. In the United States, secondary analysis as a

research strategy coincides with the rapid increase in the number of attitudinal surveys before the Second World War. The 'first notable effort' at secondary analysis from a theoretical and methodological perspective was *The American Soldier* (Glaser 1963: 11). *The American Soldier* is part of a collection of around 260 studies of several hundred thousand army personnel that was undertaken during the Second World War. Its four volumes are the accumulation of several years of investigation and include surveys and observations of soldiers' lives before, during and after combat (Lazarsfeld 1949). The work provides a fascinating insight into the lives of service personnel: for example, their relationships with their fellow soldiers and officers, their changing attitudes towards war as a result of experiencing combat and their adaptation to life as civilians. In addition, the collection is seen as a 'fine model indeed for the secondary analysis of a rich archive of data collected for other purposes' (Smith 1984: 196, cited in Williams 1989: 157). The legacy of *The American Soldier* has not just been its contribution to the field of military history. Its theoretical and conceptual developments have found influence in developing theories of class position, of race relations, of 'relative deprivation' and social adjustment; and its methodological applications have led the field in the development of attitude scaling and latent structural analysis (Glaser 1963; Lazarsfeld 1949; Williams 1989).

Other classic examples of secondary analyses include the re-analysis of the Coleman Report during the early 1970s (Mosteller and Moynihan 1972). This re-analysis was in a large part prompted by the huge interest generated by the study *Equality of Educational Opportunity* (Coleman et al. 1966). The original study found that the variation in learner outcomes in schools was strongly related to factors external to the school, such as family background and that the school itself had a relatively small part to play in mediating between different academic outcomes. Perhaps unsurprisingly, given the reception that greeted the original study, a re-analysis of the findings soon followed. Undertaken by faculty members at Harvard University, the secondary analysis identified some discrepancies, errors and issues with the primary study but largely 'affirmed and strengthened' (Smith 1972: 311) the original findings. Another example of important secondary analyses of influential studies is Elashoff and Snow's (1971) re-analysis of Rosenthal and Jacobsen's (1968) study *Pygmalion in the Classroom*.

In the UK, one important use of secondary analysis in the sociology of education is the 1980 re-analysis of the 1972 Oxford Mobility Project dataset by Halsey et al. This study is not only important for its empirical and theoretical contributions, as the subsequent book *Origins and Destinations: Family, Class and Education in Modern Britain* is one of the classics in the sociology of education.

Works such as *The American Soldier, Origins and Destinations* and *Suicide*

set the standard for secondary analysis in the social sciences, exemplified by their methodological, theoretical and substantive contributions to the development of the field. However, arguably the potential for secondary analysis as an important social science method has never fully been realised in many branches of the discipline, as many of the objections to its use (considered in Chapter 2) attest. While secondary analysis may be more established as a research method in the United States (Hakim 1982), it has remained relatively underused in many areas of the social sciences in the UK. Indeed, the 1988 edition of *The Penguin Dictionary of Sociology* contains no entry for secondary analysis. In the entry for 'official statistics' sociologists were advised to 'approach such data with caution', its use was described as 'perilous' and data collected by government for its own use considered 'meaningless' (Abercrombie et al. 1988: 170). Perhaps some minor rehabilitation of the field has occurred in the intervening years as a definition for 'secondary analysis' was suggested in the 2000 edition and the authors' objections to the use of official statistics have been somewhat mollified in these later editions (Abercrombie et al. 2000).

How widely are secondary data analysis used?

In this section, we broaden our consideration of secondary data analysis to reflect some of the methodological challenges faced by the social science community. The focus begins with a brief discussion about general methodological concerns before considering the implications these may have for the secondary analysis of numeric data.

In education, recent methodological preoccupations in both the UK and the USA have focused on the quality and relevance of research in the field. Educational research is widely viewed as having an 'awful' reputation (Kaestle 1993) of being 'not very influential, useful or well funded' (Burkhardt and Schoenfeld 2003: 3), of following fads (Slavin 1989) and of being of indifferent quality (Hargreaves 1996; Tooley with Darby 1998). While the call in both countries is for a greater unity between research and practice (Burkhardt and Schoenfeld 2003; Hargreaves 1996), there is some divergence in how this might actually be achieved. In the USA, legislation introduced in 2001 stipulates that all federally funded research must adopt scientifically based research methods (Eisenhart and Towne 2003; Olson and Viadero 2002). For some this is seen as an opportunity to elevate educational research to the status of medicine and agriculture (Slavin 2002) and for 'nurturing and reinforcing' a scientific research culture in the field (Feuer et al. 2002: 4). For others, it exemplifies the privileging of certain research methods: namely, experiments and randomised control trials, a failure to understand the

complexity of the field and a lack of commitment by the US federal government to promoting true evidence-based practice (Berliner 2002).

In the UK, general methodological concerns centre on a perceived imbalance in the types of research method adopted by educational and other social science researchers (ESRC 2006). Much of this concern is centred on the 'dubious dichotomy' (Payne et al. 2004: 153) that exists between 'quantitative' and 'qualitative' methods. For example, according to the Economic and Social Research Council (ESRC), 'the lack of quantitative skills is endemic in many areas of Social Science and ... there is an urgent need to enhance research quality' (ESRC 2006: 12). The ESRC has demonstrated its commitment to building research capacity in the field of quantitative methods through sponsorship of centres such as the Social Science Research Methods hub at Southampton University, the development of data management infrastructures such as UK Data Archive and its sponsorship of the *European Social Survey* and other internationally renowned longitudinal studies such as the *National Child Development Study* and the *British Household Panel Survey*.

In the field of education, a lack of quantitative skills among researchers is seen as being one of its most significant 'defects' (Gorard et al. 2003a: 19) and, according to key stakeholders, there is a significant lack of quantitative skills in the field:

> There is a widely acknowledged absence of quantitative research of particular kinds, especially, there's a weakness, there's a relative absence and there's no mechanism for addressing that currently. (HE researcher and ESRC Teaching and Learning Research Programme team leader, cited in Gorard et al. 2003a: 13)

In addition to concerns about a lack of quantitative work in education, there are also concerns about the quality of such work:

> I think you can get terrible quantitative work, there are people who just think there's a kind of non-problematic general linear model reality out there and you just tag variables and start with race and that's it ... switch on SPSS, press the buttons with all the defaults on is garbage in garbage out. (chief executive of research funding body, cited in Gorard et al. 2003a: 17)

Thus it can be argued that building research capacity in the use of quantitative methods is not just about increasing the volume of such research but focusing on more basic quantitative techniques. The concern appears to be not only about the relative shortage of quantitative research 'but also a shortage of the quantitative research skills required to be able to understand, and critically review, quantitative research' (Gorard et al. 2003a: 19). The next section considers these concerns in light of a review of publications submitted to the 2001 Research

Assessment Exercise (RAE) and a more recent examination of the published output of eight social science journals.

What evidence is there for a lack of quantitative skills?

In the UK, the funding councils for higher education institutions (HEIs) have highlighted the importance of comparative measures of research excellence. This has resulted in the Research Assessment Exercise (the RAE), the purpose of which is to produce quality profiles for all research activity in UK HEIs. Indeed, the RAE is the principal means by which institutions assure themselves of the quality of research undertaken in the HE sector (RAE 2008). The first RAE was carried out in 1986 with four subsequent exercises prior to the latest RAE in 2008. In *RAE 2001* research quality ratings were based on a range from 1–5*, according to how much of the work was judged to reach national or international levels of excellence. In this way, the quality ratings derived from the RAE provide a single measure of research strength in UK higher education. In this section, therefore, we report the findings from an analysis of the results for the most recently available RAE, *RAE 2001*, and examine the evidence for a lack of quantitative skills in the social sciences and in the field of education, in particular (Gorard et al. 2003b). We begin by looking at the overview reports for the four main RAE social science panels – education, social policy, social work and sociology. Their comments suggest that in some areas of the social sciences, a concern over the paucity of quantitative research does remain:

> The strengths of qualitative, and relatively small scale, research have been complemented by some increase in work that has also drawn on quantitative data. However, there is room for more approaches that use advanced quantitative methodologies and for education to play an active and significant role in methodological innovation. (Education Panel RAE 2001a)

> There was a paucity of quantitative analysis, particularly that combined with qualitative approaches. Quantitative work, while sometimes of very high quality, is rare and not always good when found. (Social Work Panel RAE 2001b)

Note that similar concerns about quantitative work in sociology and social policy and administration were not raised in the RAE 2001 overview documents for these two panels (RAE 2001b). This, of course, does not necessarily mean that such concerns do not exist (and, indeed, some of the evidence presented in the following section argues that, at least in sociology, they well might) rather it may simply be the case that they were just not reported by the *RAE 2001* panels.

Because of the variety of submissions in the field of education, the RAE panel for 2001 required that institutions also provide additional information about the methods undertaken in each submission. This theory/

method field was to be used to classify 'any conceptual framework within which the research may be set, together with a brief indication of the design and/or research method' (RAE 2001c). This means that for education, in addition to providing indicators of quality, the RAE is also able to provide a much more detailed audit of the research methods that were being used (Gorard et al. 2003b).

While the use of the theory/method field to classify research methods is not itself unproblematic, an analysis of the 8700 publications submitted to the education panel in *RAE 2001* does reveal an interesting balance of methodological approaches. Table 1.1 shows the number of times a particular method was reported as being the first or main method used in submissions to the education panel in *RAE 2001*. According to the entries reported in the theory/method field, a large number of submissions (32%) were not clearly empirical: for example, they involved literature reviews or thought pieces. Of the remainder around 14% were largely or wholly quantitative and 28% were largely or wholly qualitative – a balance of around 2:1 (Gorard et al. 2003b). Notice also the relatively lowly position of secondary data analysis, which was listed as the main method for fewer than 1% of the pieces submitted to *RAE 2001*.

The number of *actual* methods used in the sample of work submitted for *RAE 2001* can be compared with a self-reported audit of methods that was undertaken by Gorard et al. (2003b) as part of the same study and reproduced in Table 1.2. This self-audit was completed by 521 researchers who were actively engaged in empirical research in the UK in 2002–2003. They were asked to summarise their knowledge and use of a range of methods taken from a specified list of nearly 300 approaches. The approaches were collapsed into similar categories as used for the theory/method analysis described in Table 1.1. The responses indicate a considerable capacity to undertake quantitative work in the field of education: around three-quarters of the research community reported having conducted a survey and around 65% report having used a secondary numeric source of data.

If we consider the number of times researchers report using quantitative and qualitative methods in their everyday research and the types of method actually adopted in publications that were submitted to *RAE 2001*, we see that while similar proportions of researchers report using quantitative and qualitative techniques (Table 1.2), in submissions to *RAE 2001*, twice as many publications were qualitative in nature than were quantitative (Table 1.1). While this gap is perhaps smaller than expected given the stakeholders' concerns that were explored earlier, it does suggest that perhaps researchers are not quite as eclectic in their use of methods as they might see themselves to be. The differences between those *reporting* using secondary data analysis and those *actually* using this method are more apparent than for the broad area of quantitative

Table 1.1 Frequency of reported first or main method used in publications submitted to Education Panel, RAE 2001

Method	Number	%
Thought piece	1533	18
Literature review	828	10
Survey	697	8
Case study	674	8
Qualitative unspecified	494	6
Comparative	479	6
Policy study/analysis	465	5
Interview	407	5
Textual analysis	392	5
Not classifiable/no method	364	4
Historical/archive	344	4
Quantitative unspecified	271	3
Ethnomethodology	268	3
Action research	233	3
Philosophy	191	2
Observation	190	2
Programme evaluation	131	2
Longitudinal study	121	1
Linguistic/conversational analysis	102	1
Experiment	94	1
Scales/psychometry	80	1
Software for collection/analysis	65	1
Group interview	51	1
Formal tests	46	1
Diaries	37	0.4
Systematic review (meta analysis)	31	0.4
Pictures/sound	29	0.3
Intervention	25	0.3
Secondary numeric data	15	0.2
Total	**8691**	**100**

Source: Gorard et al. 2003b: 44

methods: 65% of respondents to the survey of Gorard et al. (2003b) report using the technique (Table 1.2) but it only features in the theory/method line of only 15 out of 8691 submissions to *RAE 2001* (Table 1.1).

The notion of methodological pluralism (where researchers are tolerant of a variety methods) was further examined by Payne et al. (2004) in their review of 2 years' output from four 'well-regarded mainstream or general' (p. 155) British sociology journals. As in the field of education, concerns have also been expressed in sociology about a lack of quantitative skills (Payne et al. 2004). Indeed, their analysis appears to support

Table 1.2 Frequency of actual methods used in publications submitted to Education Panel, RAE 2001

Method	Number	%
Interview	480	92
Literature review	471	90
Case study	421	81
Sampling	416	80
Observation	415	80
Textual analysis	409	79
Qualitative general	400	77
Quantitative general	393	76
Survey	391	75
Triangulation	370	71
Secondary numeric sources	340	65
Group interview	339	65
Software for collection/analysis	325	63
Pictures/sound	298	57
Evaluation	289	56
Action research	269	51
Systematic review (meta analysis)	254	49
Diaries	239	46
Experiment	212	41
Longitudinal study	199	39
Linguistic/conversational analysis	195	38
Formal tests	194	37
Scales/psychometry	184	35
Ethnomethodology	142	27
Visual/sound sources	91	17
Historical archive	69	13

N = 514 (respondents could list more than one method).

Source: Gorard et al. 2003b: 31

such claims as they found that only about one in 20 published papers used any form of quantitative analysis. Suggesting no evidence of genuine plurality of methods:

> [I]ndividual sociologists – no matter how tolerant, catholic and eclectic – are very unlikely actually to be methodological pluralists ... It is the structure of sociology that became pluralist, not sociologists themselves. (Bell and Roberts 1984: 5, cited in Payne et al. 2004)

Using RAE returns and a self-selected audit survey as a means of identifying the research skills of social scientists can only take us so far. In the next section, we shift our focus back to secondary data analysis to

examine the extent to which social researchers are as plural in their use of secondary data methods as their self-reports suggest they might be.

Secondary data analysis in practice

In an approach similar to that undertaken by Payne et al. (2004) and Gorard et al. (2003a), I have undertaken an analysis of the published output of eight mainstream and well-regarded journals in the fields of education, sociology and social work over a 7-year period. The aim of this analysis is to investigate concerns about a lack of methodological pluralism in the use of quantitative methods in these fields and, more importantly for the purpose of this book, for the use of secondary data analysis.

Three journals each in the fields of education and sociology and two in the area of social work were selected. The journals were:

Education: *British Educational Research Journal*, *Oxford Review of Education* and *Research Papers in Education*.
Sociology: *British Journal of Sociology*, *Sociology* and *The Sociological Review*.
Social Work: *British Journal of Social Work* and *International Social Work*.

All eight journals are listed in the social science citation index and were targeted at a general audience: for example *The Sociological Review* has 'a flexible approach to both its content and its style. No social topic is considered irrelevant, innovative subject matter and style are welcomed, and articles are always topical and current' (Blackwell Publishing 2005). Similarly *Oxford Review of Education* 'publishes papers on the theory and practice of education from scholars throughout the world in a variety of disciplines: philosophy, political science, economics, history, anthropology, sociology, psychology and medicine' (Taylor and Francis Group 2007). Although each of the selected journals draws contributions from around the world, in particular *International Social Work*, all had editors who were based in the UK. Because of the relatively general scope of the journals there was some degree of permeability across disciplines, for example, papers from the broad field of 'education' appeared in both the 'sociology' and 'social work' journals.

Similarly, the analysis described here does not distinguish between pieces submitted by academic departments and those from other institutions. This might make for a generous interpretation of the output of quantitative and secondary analytic pieces in the field of education where many papers which emphasised secondary analysis in particular were submitted from organisations such as the National Foundation for Education Research (NFER) and examination boards such as the University of Cambridge Local Examinations Syndicate (UCLES), rather than from HE schools of education. Additionally, as we have seen in the

Table 1.3 Use of secondary data analysis and quantitative methods, selected social science journals

Journal	Secondary data analysis	Quantitative methods	Total papers
British Educational Research Journal	34	85	274
Oxford Review of Education	30	56	220
Research Papers in Education	16	51	133
Education total	**80 (42%)**	**192 (31%)**	**627**
British Journal of Sociology	49	58	201
Sociology	26	37	294
Sociological Review	14	24	211
Sociology total	**89 (75%)**	**119 (17%)**	**706**
British Journal of Social Work	15	95	422
International Social Work	18	86	261
Social work total	**33 (18%)**	**181 (27%)**	**683**
All journals	**202 (41%)**	**492 (24%)**	**2016**

introductory sections of this chapter, there is no real clear and unambiguous definition of secondary data analysis. I have, therefore, taken a very inclusive approach to deciding whether or not articles adopt secondary analytic techniques. Selection criteria were limited to numeric empirical pieces which reported aggregate secondary data, as well as those which undertook an analysis of raw data. Bearing these caveats to a rather imprecise art in mind, we now consider the results of the journal search, beginning with the frequency of pieces which adopted quantitative and/or secondary data analysis methods (Table 1.3).

About one-quarter of all the papers that were reviewed adopted some form of quantitative method (492/2016), of these around 41% (202/492) used secondary data analysis. Overall, fewer than 10% of papers (202/2016) involved the analysis of secondary data. The findings for the individual subject areas are considered next.

Sociology

The findings for the 'sociology' journals reflect some of the concerns of Payne et al. (2004) in that only 17% of papers published were quantitative in nature. However, these overwhelmingly used secondary data analysis: 89 out of 119 numeric pieces in the 'sociology' journals adopted some form of secondary data analysis. The range of secondary sources

used in 'sociology' journals was extensive and included the UK population census and government-sponsored surveys such as the *Labour Force Study*, among many others. Examples of the range of secondary sources used in the field of sociology include data from the *European Values Survey* to examine secular beliefs in Europe (Halman and Draulans 2006), childhood poverty and early parenthood explored through the *National Child Development Study* (Hobcraft and Kiernan 2001) and the *British Family Resources Survey* being used to examine the pension prospects of minority ethnic groups (Ginn and Arber 2001). The use of international secondary sources was also apparent. For example, the national mobility surveys in Japan and Israel were used in order to study the relationship between class structure and social mobility among women and ethnic groups in these two countries respectively (Shirahase 2001; Yaish 2001).

Education

Around one-third of the papers published in the three 'education' journals involved some use of quantitative methods. The range of quantitative approaches varied from relatively small-scale questionnaire surveys, sometimes combined with other approaches (for example, Canning 2000; Edwards and Protheroe 2003) to large-scale longitudinal national cohort studies (for example, Driessen et al. 2005). However, in contrast to the 'sociology' journals, the papers that used some component of secondary data analysis comprised a relatively small sub-section of the quantitative genre. Here only 80 of the 192 numeric pieces used secondary data analysis (see Table 1.3). In the 'education' journals that were reviewed here, the majority of pieces that included any analysis of secondary data were largely based on the analysis of school examination data (for example, Goldstein 2001; Strand 2002). Much less use was made of secondary data from other sources. However, there were some exceptions. For example, Payne's use of *Youth Cohort Study* data to look at the impact of part-time work among year 12 and year 13 students (Payne 2003), the use of the *1958 British Birth Cohort Study* to examine the relationship between family background and school subject choice (Van de Werfhorst et al. 2003) and the analysis of participation data from the *Higher Education Statistics Agency* in an examination of patterns of inclusion among disabled students in higher education (Riddell et al. 2005). A small number of papers also used secondary data from international sources. For example, Post's use of the *Hong Kong National Census* to look at educational stratification and access to higher education (Post 2003) and use by Williams et al. of the *Programme for International Student Assessment (PISA)* data to model achievement among 15 year olds in OECD countries (Williams et al. 2005).

Social work

In common with 'education', less than one-third of publications in the 'social work' journals had a numeric component. As in 'education' a large proportion of these papers reported the findings of relatively small-scale primary surveys, for example Mosek and Adler (2001) and Wardell et al. (2000). Fewer than 20% (33 out of 181) of these numeric pieces involved the analysis or the presentation of secondary data. One reason for this may be the relative paucity of good-quality secondary datasets covering contemporary research areas of interest in the field. However, as the wide range of topics covered in the 'social work' journals attests, social work, like education, is a large multidisciplinary field. Indeed, that there are opportunities in the field of social work for secondary analysis is exemplified by the range of datasets that were reported in the journals included in this study. These include Buchanan and Flouri's (2001) use of the *National Child Development Study* to examine the relationship between parental background and familial emotional support in adulthood and Bertera's (2003) use of United Nation's data on life expectancy to explore social services for the aged in Cuba. There were also studies that combined a range of secondary sources, such as Ji's examination of the effects of risk factors on rates of homelessness in the USA which used secondary data from the US Census Bureau, the National Low Income Housing Coalition, the Department of Housing and Urban Development, the Urban Institute and the Interagency Council on the Homeless (Ji 2006).

This quick analysis of the frequency of use of numeric and secondary analytic techniques in three areas of the social sciences to some extent reinforces the view of the Gorard et al. (2003a) stakeholders that quantitative methods are underused in social science research, although the finding that around one-third of the papers in the 'education' journals used numeric methods perhaps gives some cause for optimism. However, in the 'sociology' journals, where fewer than 20% of papers adopted quantitative techniques, the concerns of Payne et al. (2004) about a lack of methodological pluralism appear to remain true. With regard to secondary data analysis, although a relatively large proportion of papers in the 'sociology' journals adopted this approach, the same cannot be said of the 'social work' and 'education' journals. Indeed, in the field of education, although the results of the Gorard et al. survey (2003a) might suggest that UK education researchers report using secondary data analysis, relatively few actually use the technique in their research – further evidence perhaps of a lack of methodological pluralism.

Summary

This first chapter has sought to introduce the reader to the field of secondary data analysis. It is a field with a long pedigree that extends back to the pioneers of the Victorian social surveys through to the opinion polls, censuses, administrative records and international surveys of today. Despite the huge potential of secondary data analysis methods, there is limited consensus on a definition for the term, although commentators tend to agree that it involves some form of re-analysis or reporting of existing data. Such flexibility in its definition perhaps reinforces its suitability to involve any analysis involving the re-interpretation of existing data which bring new methodological and theoretical perspectives or which adopt the original or novel research questions and which are undertaken by the original researcher or individuals new to the data. It is also important to emphasise that the analysis of secondary data need not involve the manipulation of complex and large datasets. Indeed, much of the data that are available for secondary analysis are in the form of aggregate data – that is, data that have already been analysed and are presented in summary form.

The second part of this chapter considered the extent to which secondary data analysis is currently used in social research. It took as its starting point contemporary concerns that quantitative methods are underused in the field and, by focusing specifically on the field of education, examined the disparity between the type of methods researchers claim to use and their actual published output, which indicates what methods they actually do use; an examination that points, in particular, to a lack of methodological pluralism in the use of secondary data in education research. One of the reasons why secondary data analysis is relatively underused in social research can perhaps be attributed to the wide range of criticisms and concerns that the field attracts. It is these potential pitfalls of secondary analysis that form the focus of the next chapter.