

# Reworking Qualitative Data

## What is Secondary Analysis?

Contributors: Janet Heaton  
Editors: Janet Heaton  
Book Title: Reworking Qualitative Data  
Chapter Title: "What is Secondary Analysis?"  
Pub. Date: 2004  
Access Date: October 17, 2013  
Publishing Company: SAGE Publications Ltd  
City: London  
Print ISBN: 9780761971436  
Online ISBN: 9781849209878  
DOI: <http://dx.doi.org/10.4135/9781849209878.n1>  
Print pages: 1-19

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781849209878.n1>

[p. 1 ↓]

# What is Secondary Analysis?

The analysis of pre-existing data in social research [2](#)

Functions of secondary analysis [8](#)

Modes of secondary analysis [12](#)

Conclusion [15](#)

Notes [16](#)

Secondary analysis is best known as a methodology for doing research using pre-existing statistical data. Social scientists in North America and Europe have been making use of this type of data throughout the twentieth century, although it was not until 1972 that the first major text on the research strategy, *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities* by Herbert H. Hyman, was published. Since then, the literature on secondary analysis of quantitative data has grown considerably as the availability and use of these data has expanded. There is now a substantial body of work exploring different aspects of the methodology, including several textbooks describing the availability of statistical data sets and how they can be used for secondary research purposes (see Dale et al., 1988; Hakim, 1982; Kiecolt and Nathan, 1985; Stewart and Kamins, 1993), as well as critical commentaries on the scientific, ethical and legal aspects of sharing this type of data in the social sciences (see Fienberg et al., 1985; Hedrick, 1988; Sieber, 1988, 1991; Stanley and Stanley, 1988).<sup>1</sup> Accordingly, the terms 'secondary analysis', 'secondary data' and 'data sharing' have become synonymous with the re-use of statistical data sets.

However, in recent years interest has grown in the possibility of re-using data from qualitative studies. Since the mid-1990s a number of publications have appeared on the topic written by researchers who have carried out ground-breaking secondary

analyses of qualitative data (Heaton, 1998; [p. 2 ↓ ] Hinds et al., 1997; Mauthner et al., 1998; Szabo and Strang, 1997; Thompson, 2000a; Thorne, 1994, 1998), by archivists involved in the preservation of qualitative data sets for possible secondary analysis (see Corti et al., 1995; Corti and Thompson, 1998; Fink, 2000; James and Sorensen, 2000) and by academics interested in these developments (see Alderson, 1998; Hammersley, 1997a; Hood-Williams and Harrison, 1998). The extension of secondary analysis to qualitative data raises a number of questions about the nature of this research strategy. What is secondary analysis? How does the secondary analysis of qualitative data compare to that of quantitative data? And how is secondary analysis distinct from other quantitative and qualitative methodologies used in social research?

In this chapter, I explore these issues by comparing and contrasting the ways in which secondary analysis has been conceptualized as a methodology for re-using quantitative and qualitative data. I begin by examining the types of pre-existing data which are deemed to be subject to secondary analysis and related methodologies. This is followed by an outline of the purported functions of secondary analysis and how these are more-or less - accepted in relation to the re-use of quantitative and qualitative data. In the last section, I describe three modes of secondary analysis and the emphasis given to each in existing conceptualizations of quantitative and qualitative secondary analysis respectively. The chapter concludes with a provisional definition of secondary analysis which clarifies the focus of this book, as well as providing a basis for the subsequent exploration of the theoretical and substantive issues arising from this conceptual overhaul.

## The analysis of pre-existing data in social research

Although the secondary analysis of quantitative data is an established methodology in social research, there is no single, unequivocal definition of the approach. Existing conceptualizations of quantitative and qualitative secondary analysis tend to refer to various propositions, some of which are more accepted than others. The first and most rudimentary principle of secondary analysis is that it involves the use of *pre-existing* data. This view is exemplified by the following who, writing on the secondary analysis

of statistical data, claim that: ‘Secondary analysis must, by definition, be an empirical exercise carried out on data that has already been gathered or compiled in some way’ (Dale et al., 1988: 3). There are, however, notable differences in the types of pre-existing data which are subject to secondary analysis, depending on the nature and origin of the material.

[p. 3 ↓ ]

Table 1.1 Examples of pre-existing quantitative data used in social research

Type	Examples
Census data	United States since 1790. England and Wales since 1801
Institutions' administrative data	Hospital records; employee records
Public records	Public Record Office; libraries
Social surveys (including omnibus and uni-purpose surveys by research centres and governments)	General Social Survey (GSS) in United States. British Social Attitudes Survey (BSAS) and General Household Survey (GHS) in Britain
Longitudinal studies	Terman database 1922 – present in United States. Longitudinal Study (LS); National Child Development Study (NCDS) and National Survey of Health and Development in Britain

## Quantitative data

Pre-existing data used in quantitative secondary analysis has been derived from various activities, including research projects carried out by academics, government agencies and commercial groups, as well as the administrative work of public authorities and other organizations that routinely keep records for management purposes.<sup>2</sup> Table 1.1 shows examples of the kinds of statistical data that have been used in secondary research in the social sciences, derived from previous research studies and from other contexts.<sup>3</sup>

For early social researchers, the main types of pre-existing quantitative data available to them were census and administrative records. Durkheim (1952), for example, used both types of data in his classic study on suicide and Miller (1967) used administrative data in her study of former ‘mental patients’ adjustment to the world after discharge from hospital. As Hyman (1972) observes these resources were more difficult to use in the past. Nowadays census data can be stored, shared and analysed with the help of computers; services have also been developed to help distribute the data and to advise

researchers on how to use it. Unsurprisingly, the use of census data has expanded, while administrative data is still used on a more occasional basis. It has been suggested that more use could be made of the latter data in areas such as nursing research (see Herron, 1989; Jacobson et al., 1993; Lobo, 1986; Reed, 1992; Woods, 1988).<sup>4</sup>

Since the 1960s, various types of social surveys have been conducted by government agencies, academics and commercial organizations and the [p. 4 ↓] resulting data made available for secondary research purposes. Hakim (1982) distinguishes multi-purpose (or omnibus) social surveys from those designed with an exclusive primary focus. As the name suggests, multi-purpose surveys are carried out in order to provide data for multiple uses (and users). In the United Kingdom, examples of multi-purpose surveys include the General Household Survey (GHS) and Labour Force Survey (LFS), which are carried out by the Office for National Statistics (ONS). The British Social Attitudes Survey (BSA) conducted by the National Centre for Social Research (NatGen) is another example of a survey carried out to provide data for others to analyze. Its American equivalent, the General Social Survey (GSS), has been carried out annually since 1971 by the National Opinion Research Center (NORC) based at the University of Chicago. Cross-national projects have also been developed, such as the International Social Survey Programme (ISSP) (Procter, 1995).

In contrast, more exclusive social surveys are designed to investigate particular research questions and are conducted on a one-off or less regular basis. Given that such surveys are designed to examine specific issues, the scope for secondary analysis may be more restricted compared to data derived from omnibus social surveys (Dale et al., 1988: 9). The same caveat applies to other statistical data collected for specific rather than generic research purposes. However, these data sets are still used as a resource for secondary research purposes. For example, data from the Nuffield study of social mobility in England and Wales (Goldthorpe, 1980) have been re-used on various occasions.

Some statistical data sets have been generated with a view to both primary and secondary research uses. For instance, longitudinal studies follow up a population cohort over time, collecting data on topics of long-term interest and further topics for investigation may be introduced as the study progresses. These data are subject

to primary analysis by the researchers who collected the data to address particular research questions; they may also be archived and used by other researchers to address additional research questions. Longitudinal studies have been most extensively carried out in the United States. The oldest ongoing American longitudinal project is the Lewis Terman study which began in 1921/2. It was originally designed to investigate the maintenance of early intellectual superiority over a ten-year period, and subsequently expanded to examine the life paths into adulthood of gifted individuals and the experiences of these individuals and their families. Data from the Terman database have been used in secondary studies, for example, Elder et al. (1993) used its data on World War II veterans in their secondary study of military service in World War II and its effect on adult development and aging.

[p. 5 ↓ ]

Table 1.2 Examples of pre-existing qualitative data used in social research

Type	Examples	Methodology
Non-naturalistic or artefactual data (solicited for research studies)	Fieldnotes Observational records Interviews Focus groups Questionnaires (responses to open-ended questions) Diaries (solicited) Life stories	Secondary analysis
Naturalistic data (found or collected with minimal interference by researchers)	Life stories Autobiographies Diaries (found) Letters Official documents Photographs Film Social interaction	Documentary analysis      Conversation analysis

Longitudinal studies have also been carried out in other countries, though not to the same extent. Hakim (1982) describes five national longitudinal studies carried out in England and Wales or Great Britain, of which only three followed up the sample beyond two years: the OPCS Longitudinal Study (LS), the National Child Development Study (NCDS) and the National Survey of Health and Development (NSHD). Other nonnational longitudinal projects carried out in England include the Newsons' study of child development (Newson and Newson, 1963, 1968, 1975). New 'linked' data sets may also be generated by combining selected data from these resources (Dale et al., 1988). For example, the aforementioned OPCS Longitudinal Study, now known as the ONS Longitudinal Study (LS), links data obtained from the 1971 Census onwards with data obtained annually from public records for those born on one of four days of the year.

## Qualitative data

In contrast to quantitative research where, as we have seen, secondary analysis encompasses the use of pre-existing data derived from research and other contexts, in qualitative research different methodologies have been used for the analysis of 'non-naturalistic' data that were solicited by researchers, and 'naturalistic' data that were 'found' or collected with minimal structuring by researchers (see Table 1.2).<sup>5 6</sup>

[p. 6 ↓ ]

In qualitative research secondary analysis is more narrowly conceptualized as a methodology for the study of non-naturalistic or artefactual data derived from previous studies, such as fieldnotes, observational records, and tapes and transcripts of interviews and focus groups. However, unlike quantitative research, there is no tradition of re-using data from previous qualitative studies. On the occasions where research data have been re-used it has tended to be in exceptional circumstances, such as when anthropologists have analysed the fieldnotes of researchers who have died before completing their work and published the results posthumously (Sanjek, 1990a). It is only relatively recently that researchers have begun to recognize the potential for re-using the various types of qualitative data produced in the course of social research, and to publish secondary studies using these resources (Heaton, 1998).

By contrast, naturalistic data such as diaries, essays and notes, autobiographies, dreams and self-observation, photographs and film have traditionally been studied using the more established methodology of documentary analysis (Jupp and Norris, 1993; Plummer, 1983, 2001; Scott, 1990). Documentary analysis has also been used to study some types of textual and non-textual data from qualitative studies which are near-naturalistic in that they were obtained with minimal shaping by researchers (as in unstructured interviews), or by using unobtrusive or even covert methods. However, some types of qualitative data can be construed as naturalistic or non-naturalistic, depending on how they originated. For example, diaries can be kept as a personal record and later 'found' and examined by researchers; alternatively, they can be designed and completed as a research tool. Similarly, life stories may be told

and recorded in autobiographies, biographies or interviews, and may be more or less structured by journalists, biographers or researchers. Thus, although secondary analysis and documentary analysis tend to focus on non-naturalistic and naturalistic data respectively, this distinction is not always clear-cut and hence there is some overlap in the types of data which are subject to these methodologies.

Life stories solicited for qualitative studies are unique data in that although they tend to be collected primarily for single use, as part of a research study investigating a specific question, they are also recorded with the intention of archiving them for possible future use in other research. For this reason they are similar to the kind of statistical data obtained in longitudinal studies in terms of being collected with both primary and secondary uses (and users) in mind. However, although life stories have been recognized as potentially valuable secondary resources (Atkinson, 1998; Thompson, 1998) and probably archived more than any other varieties of qualitative data, it is not clear to what extent, and how, [p. 7 ↓] researchers have used these pre-existing data. There is little published national or international information on the extent and nature of use of these resources generally.<sup>7</sup> In addition, there have been no recent reviews of the use of life stories in social research and whether these were obtained from archives or through primary research.<sup>8</sup> There is also little guidance on how to do social research using pre-existing life stories: textbooks tend to focus on how to collect, record and analyze life stories in the context of primary research rather than how to access and undertake secondary research using pre-existing life stories deposited in archives (see Atkinson, 1998; Langness, 1965; Miller, 2000; Yow, 1994).

The nearest qualitative equivalent to omnibus social surveys designed to provide statistical data for use in multiple studies are projects which have been undertaken to collect and preserve textual and non-textual remnants of social life. The Mass Observation project is a prime example of this type of work. Described as an 'anthropology of ourselves' (quoted in Calder and Sheridan, 1984: 247), various materials were gathered by the investigators on the project between the 1930s and 1950s and more recently during the 1980s and 1990s. Various non-naturalistic and naturalistic data were collected, including reports of interviews and observations, time sheets and diaries, and original documents, such as press cuttings, pamphlets, and photographs. These data were collected and archived with a view to their value as



a historical resource and more general utility. The archivist (Sheridan, 2000) notes that these data have been used by academic researchers from a range of disciplines (including art history, social history, anthropology, psychology, sociology, media and cultural studies) as well as by the wider community (including the media, novelists, artists, teachers and students). She adds that since 1970 social research using the data has taken four main forms: as evidence in historical research; in the study of the Mass Observation movement itself; in methodological research; and to inform the development of related initiatives. In principle, the re-use of some of the data from the Mass Observation project may therefore be defined as documentary or secondary analysis, depending on which type of data is used and for what purposes (of which more below).

Another methodology which has been used in social research to study naturalistic data in the form of recordings of everyday social interaction is conversation analysis (see Atkinson and Heritage, 1984; Hutchby and Woofit, 1998; ten Have, 1999). These data include tapes and transcripts of audio and/or visual recordings made and transcribed by the researchers personally, or by other researchers. Indeed, conversation analysts often develop analyses based on pre-existing data that have been shared within this disciplinary network. Lengthy extracts of the data, transcribed using [p. 8 ↓] a detailed system of annotating verbal and non-verbal interaction, are also reproduced in published studies in order to allow for independent access to, and scrutiny of, the data upon which the analyses are based. However, despite these practices, no distinction is made between 'primary' and 'secondary' conversation analysis. Rather, these naturalistic data are seen as pristine and in no way an artefact of the research process; as such, they are assumed to be open to analysis by all on an equal basis, regardless of who collected the data and rendered it for analysis. Conversation analysis is therefore an interesting example of a methodology that involves the re-use of qualitative data but which is not regarded by exponents as a secondary methodology because of the perceived unadulterated and naturalistic nature of the data concerned.

# Functions of secondary analysis

It has been suggested that quantitative and qualitative secondary analysis can be used for three broad purposes although, as I show below, some of these claims are a matter of debate.

## Investigation of new or additional research questions

First, it has been proposed that secondary analysis allows researchers to put to new or additional uses data that were originally collected for other research purposes or, in the case of quantitative data, administrative purposes. This is illustrated by the following quotations, in which the secondary analysis of quantitative data is defined as:

...the study of specific problems through analysis of existing data which were originally collected for other purposes. (Lipset and Bendix, 1959. Quoted in Glaser, 1962: 71)

...the *extraction of knowledge on topics other than those which were the focus of original surveys*. (Hyman, 1972: 1. Original emphasis)

Any further analysis of an existing data set which presents interpretations, conclusions, or knowledge additional to, or different from, those presented in the first report on the inquiry as a whole and its main results. (Hakim, 1982: 1)

...[the study of] a problem by analysing data that originally were collected for another study with a different purpose. (Woods, 1988: 334)

Likewise, it has been suggested that the secondary analysis of qualitative data involves: **[p. 9 ↓ ]** the use of an existing data set to find

answers to a research question that differs from the question asked in the original or primary study. (Hinds et al., 1997: 408)

Thus, quantitative and qualitative secondary analysis have each been conceptualized as methodologies for conducting free-standing studies using pre-existing data originally collected for other purposes.

Although this proposition is generally accepted, it has been pointed out that this way of defining secondary analysis is limited in that it fails to acknowledge that some projects are actually designed to supply data for sundry secondary studies (Dale et al., 1988). The data from omnibus surveys are, for example, collected and prepared for secondary research purposes and are not subject to primary analysis as such. To a lesser extent, the same applies to data from longitudinal studies which are designed both to address particular primary research questions and to supply long-term data for future secondary research.

## Verification, refutation and refinement of existing research

Second, and more controversially, it has been suggested that secondary analysis can be used as a means of verifying, refuting or refining the findings of primary studies through the re-analysis of data sets. This proposition is illustrated by the following quotations, where it is suggested that the secondary analysis of quantitative data involves the:

... re-analysis of data for the purpose of answering the original research question with better statistical techniques, or answering new questions with old data. (Glass, 1976: 3)

... re-analysis of data originally collected and analysed by another investigator addressing the same question, a different question, or applying different methods of analysis. (Woods, 1988: 334)

... use of data gathered for the purposes of a primary research analysis (original research) but looks at questions not addressed by the original investigator, or addresses the same questions using different methods of analysis. In large-scale studies, secondary analysis is further used by investigators to validate the results from the primary analysis; that is, re-analysis and testing of new hypotheses may support further or dispel initial findings. (McArt and McDougal, 1985: 54)

This principle is generally accepted in quantitative secondary analysis, although it too has been queried. For example, Hyman (1972: 76, footnote 3) argues that the re-analysis of survey data for the purpose of verification is not a type of secondary analysis on the grounds that 'the focus [of re-analysis] is on the original problem that the survey was intended to illuminate' and not on a new topic. It could be countered, however, that [p. 10 ↓ ] re-analysis nevertheless has the potential to generate additional knowledge and insights through the production of findings which may or may not corroborate the primary work. In addition, where such studies are carried out by independent researchers, re-analyses are also secondary to, and free-standing of, the original research in the sense of being carried out by other analysts, even though each study investigates the same questions using the same data.<sup>9</sup>

To a lesser extent, secondary analysis has also been suggested as a possible means of re-investigating questions addressed in previous qualitative studies (see Corti, 2000; Szabo and Strang, 1997). For example, Szabo and Strang (1997: 67) have claimed that it can involve looking at 'the same questions with different analysis methods'. However, as Hammersley (1997a) has pointed out, whether the findings of qualitative research can be verified in the same ways as quantitative research is debatable, given that there are different philosophical perspectives on the respective nature of quantitative and qualitative data and associated research methods.<sup>10</sup> Qualidata have also identified resistance to the re-use of qualitative data for this purpose (Corti, 1999; Corti et al., 1995). This proposition appears therefore to be more contentious in relation to qualitative secondary analysis, reflecting wider epistemological tensions in quantitative and qualitative research which will be explored in later chapters.

# Synthesis of research

The last and most contentious of the purported functions of secondary analysis is whether or not it encompasses various types of meta-research designed to synthesize knowledge arising from existing studies. Meta-analysis and related techniques for examining evidence from quantitative research have been described by some commentators as forms of secondary analysis (Jensen and Allen, 1996; Woods, 1988). However, others have stressed that these various meta-research strategies are distinct from secondary analysis in that they seek to identify, appraise, and aggregate or synthesize existing knowledge on a particular topic. Thus, Glass distinguishes meta-analysis from secondary analysis, describing the former as the 'analysis of analyses' (1976: 3). Kielcolt and Nathan also claim that meta-analysis is unique in that it seeks to integrate the findings from a 'universe' or 'sample' of investigations of some phenomenon: 'that is, the study itself becomes the unit of analysis' (1985: 10).

Similarly, some commentators have claimed that qualitative versions of meta-analysis are forms of secondary analysis. For example, 'aggregated analysis', which involves the synthesis of findings from several qualitative studies, has been described as a variety of secondary analysis [p. 11 ↓ ] (Estabrooks et al., 1994). However, in contrast, Thorne has argued that 'whereas meta-analysis serves as a strategy for synthesis of research findings, secondary analysis provides a mechanism for extending the contexts in which we are able to use and interpret qualitative research data' (1998: 548). In addition, I have previously suggested that meta-analysis, metasynthesis, meta-studies, systematic reviews, narrative analysis and literature reviews differ from secondary analysis on two grounds. First, the former research strategies are concerned with appraising and summarizing existing knowledge, and not with exploring new research questions or verifying the results of individual studies. Secondly, they mainly involve the study of research reports, seldom reverting to the raw data itself (Heaton, 1998). For these reasons, I regard quantitative and qualitative meta-research strategies as separate methodologies from secondary analysis.<sup>11</sup>

However, some forms of meta-research are more difficult to distinguish from secondary analysis than others. For example, in quantitative research, there is an approach

involving so-called 'individual' meta-analysis in which researchers do revert to the raw data in order to standardize analyses across studies and/or to re-examine the data in new ways (Smith and Egger, 1998). This may be necessary for the purposes of synthesizing results when they are not reported in a standard way, or cannot be disaggregated, allowing findings relating to particular groups to be scrutinized. While this procedure is consistent with the meta-analyst's aim of synthesizing existing findings, the researcher is also, in effect, adding to what was previously known through the re-modelling and re-analysis of the micro data. In producing new knowledge, the boundaries of this form of meta-analysis therefore blur with those of the secondary analysis of data sets.<sup>12</sup>

Similarly, in qualitative research, Noblit and Hare contend that meta-ethnography: 'can be considered a complete study in itself. It compares and analyzes texts, creating new interpretations in the process. It is much more than we usually mean by a literature review' (Noblit and Hare, 1988: 9). In 'creating new interpretations' this approach appears to overlap with secondary analysis. Conversely, it is debatable whether 'aggregated analysis' is, as the authors (Estabrooks et al., 1994) claim, a variety of secondary analysis or in fact a meta-research strategy. On the one hand, theory development is the primary aim of aggregated analysis, not merely the synthesis and comparison of findings; in this respect it resembles secondary analysis. On the other hand, in aggregated analysis the researcher does not return to the original data. Instead, the findings reported by the original researcher are used - although the authors note that, unlike meta-ethnography (and presumably other meta-research strategies), aggregated analysis involves 'interpretative techniques' and is not a 'context-stripping [p. 12 ↓] activity' (Estabrooks et al., 1994: 505). Thus, aggregated analysis appears to be a hybrid methodology, drawing on aspects of secondary analysis and meta-analysis in its approach to social research using pre-existing data.

Table 1.3 Modes of secondary analysis

Mode	Source of data	Data donors
Formal data sharing (via intermediary services)	General data archives Special collections Commercial companies Public records Publications	Primary researchers for use in independent secondary studies Organizations for use in independent secondary studies
Informal data sharing (by special request and private networks)	Primary researchers' personal data collections Disciplinary networks Organizations' in-house records	Primary researchers for use in secondary studies by independent researchers Primary researchers for use in secondary studies by independent researchers and the data donors Organizations for non in-house secondary use
Personal or inside secondary analysis	Primary researchers' personal collections Organizations' in-house records	Primary researchers for use in personal secondary analyses of their own data Organizations for secondary use in-house

## Modes of secondary analysis

There are three main modes of secondary analysis (see Table 1.3). In the first - 'formal' data sharing - secondary analysis is carried out using data sets that have been officially made available for data sharing (although access may be controlled or restricted). This includes data sets deposited in general data archives and special collections, and those managed by commercial companies, as well as raw data published in research reports or other media. In this mode of secondary analysis, the data will have been collected and deposited by one group of researchers or organizations (data donors) and accessed by another (data users), hence secondary studies using these resources are carried out by independent researchers.

This is not necessarily the case in 'informal' data sharing where data is either obtained directly from primary researchers and organizations by request, or indirectly through private disciplinary networks (as in [p. 13 ↓ ] conversation analysis). In this mode of data sharing, the secondary analysts may or may not invite the primary researchers who donated the data to be part of the research team carrying out the secondary research. Two or more primary researchers may also combine or pool data sets from their previous work and jointly analyze them as part of a new secondary project; in this case, as primary-cum-secondary analysts, the researchers will be working with a mix of independently and self-collected data sets.

In contrast, the last mode of secondary analysis does not involve any data sharing. Instead, researchers re-use their own data, or what I referred to as 'auto-data' in Heaton (2000), and organizations carry out internal secondary analyses of their own records.

Personal or inside secondary analysis is unique in that it is carried out by the same researchers and organizations that originally compiled the data, and no one else.

In existing conceptualizations of the secondary analysis of quantitative data it is often assumed that such studies are carried out using data which have been collected by other researchers,<sup>13</sup> although whether this is via formal or informal data sharing is not usually specified. This is illustrated by several of the above quotations, as well as the following, in which the methodology is described as:

... any analysis of data collected originally by persons other than the analyst. (Miller, 1982: 719)

... a form of research in which the data collected by one researcher are reanalysed by another investigator, usually to test new hypotheses. (Polit and Hungler, 1983. Quoted in McArt and McDougal, 1985: 54)

... the application of creative analytical techniques to data that have been amassed by others. (Kiecolt and Nathan, 1985: 10)

... the further analysis of data by anyone other than those responsible for its original commissioning or collection. (Dale et al., 1988: 4)

However, Hyman (1972) has shown that the secondary analysis of quantitative data has not always been the prerogative of independent analysts. In discussing the secondary analysis of survey data he distinguishes pure from what he calls 'semisecundary' analysis. While the former is exemplified by analysts who make use of survey data deposited in data archives, the latter refers to three other types of secondary analysis, one of which is performed by the original, primary investigators. According to Hyman, this approach was in practice before the era of data banks, when researchers drew upon the wealth of data that they or their organizations had amassed themselves.

While the re-use of auto-data has been largely disregarded in the more recent literature on the secondary analysis of quantitative data as attention [p. 14 ↓] has shifted to the use of major, independently collected data sets, this and other modes of secondary analysis have all been recognized in work on the re-use of qualitative data sets. For



instance, Thorne (1994, 1998) envisaged at least five discrete types of qualitative secondary analysis, involving both the re-use of other researchers' data sets and researchers' own data from their previous primary work. In the first variety, called 'analytic expansion', researchers make further use of their own data 'to answer new or extended questions' (Thorne, 1998: 548). In 'retrospective interpretation', researchers examine new questions which were raised but not addressed in the context of the primary study (although here Thorne does not specify whether this applies to auto-data sets and/or independently collected data). The third type of secondary analysis is used in relation to other researchers' data sets, which are subject to 'armchair induction'. Using this approach, researchers apply inductive methods of textual analysis for the purposes of theory development. A fourth and additional approach to theory development using secondary analysis is that of 'amplified sampling'. This involves the comparison of several distinct and theoretically representative data sets. The fifth and final approach outlined by Thorne is 'cross-validation'. Again this involves the use of pre-existing, independently collected data sets, this time in order to 'confirm or discount new findings and suggest patterns beyond the scope of the sample in which the researcher personally has been immersed' (Thorne, 1994: 267).

Other authors have also examined the potential of doing qualitative secondary analysis using these different sources of qualitative data. For example, Szabo and Strang (1997) and Mauthner, et al. (1998) discuss their respective experiences of re-using qualitative data that they personally had previously collected and the issues raised by this mode of secondary analysis. In addition, West and Oldfather (1995) have outlined a methodology called 'pooled case comparison' which is based on the informal sharing of qualitative data. Defined as a way of 'comparing separate but similar studies *ex post facto*', it involves the use of raw data from two primary studies while setting aside any 'categories and properties from previous analyses' (West and Oldfather, 1995: 454). The approach is similar to the 'amplified sampling' form of secondary analysis outlined by Thorne (see above), in which comparisons across two or more independently collected data sets are made, although just two data sets are used, each of which were originally collected by one or other of the secondary analysts. Finally, Qualidata and other archivists have highlighted the potential of re-using qualitative data sets which have been formally archived and made available for data sharing (see Corti et al., 1995; Corti and Thompson, 1998; James and Sorenson, 2000).

[p. 15 ↓ ]

## Conclusion

'Secondary analysis' is a rather nebulous concept. It has been variously associated with different types of quantitative and qualitative data, different functions, and different *modi operandi*. Conceptually, there are similarities in the ways in which quantitative and qualitative secondary analysis have been defined, and some overlap between these approaches and other methodologies, such as meta-analysis and documentary analysis. However, there are also some notable variations in these concepts which suggest that quantitative and qualitative secondary analysis are not equivalent research strategies, and that these methodologies can be broadly distinguished from related research strategies.

Three key variations in the conceptualizations of quantitative and qualitative secondary analysis have been observed. The first concerns the types of pre-existing data which are deemed to be subject to secondary analysis. While quantitative secondary analysis encompasses the re-use of data derived from research and other contexts, qualitative secondary analysis has a narrower focus on non-naturalistic data derived from previous qualitative studies. Second, while it is generally accepted that quantitative and qualitative secondary analysis can be used to investigate new or additional research questions, it is a matter of contention as to whether secondary analysis can be used to verify previous qualitative research in the same ways as it is in quantitative research. Finally, another difference in quantitative and qualitative secondary analysis concerns the sources of data which may be used. In quantitative secondary analysis, it is generally assumed that it involves the use of other researchers' data, obtained via formal or informal data sharing. However, in qualitative secondary analysis, the potential of re-using one's own data has also been recognized alongside these modes of secondary analysis.

The focus of qualitative secondary analysis on non-naturalistic data derived from previous studies distinguishes it from other qualitative methodologies, such as documentary analysis and conversation analysis, that are used to study more naturalistic pre-existing qualitative data which are not an artefact of research work.

However, as we have seen, some qualitative data, such as life stories, can be construed as naturalistic or non-naturalistic depending on how they originated and hence there is, in principle, some overlap between the methodologies used to examine these types of data. Similarly, while there is in principle some overlap between secondary analysis and quantitative and qualitative varieties of meta-analysis, I have suggested that they are broadly distinguishable in terms of the functions of these approaches and the nature of the pre-existing material that is involved. Thus, secondary analysis is not [p. 16 ↓ ] regarded as a methodology for the synthesis of previous research, but as a methodology for investigating new research questions or for verifying previous research. However, it is recognized that the latter function is also a matter of contention in relation to qualitative secondary analysis in particular.

These findings raise issues about whether existing conceptualizations of quantitative and qualitative secondary analysis are reflected in practice. For example, it is one thing to claim that secondary analysis may be used to verify research but it is another as to whether it has been used for this purpose in the qualitative secondary studies conducted to date.<sup>14</sup> Similarly, it would be interesting to know to what extent researchers have used their own or other researchers' qualitative data in the secondary studies carried out to date. In order to further investigate the ways in which secondary analysis has been developed and actually used as a qualitative methodology to date, the following, provisional, definition of secondary analysis was adopted as the basis for an empirical review of existing qualitative secondary studies:

*Secondary analysis is a research strategy which makes use of preexisting quantitative data or pre-existing qualitative research data for the purposes of investigating new questions or verifying previous studies.*

This formulation qualifies the type of qualitative data involved (limited to data from research studies) and recognizes two of the proposed functions of secondary analysis (but not the synthesis of research findings). At the same time, it leaves open the issues of whether or not qualitative secondary analysis can be used to verify the results of previous work, and which sources of qualitative data are used. It should be stressed that this is a working definition, the adequacy of which will be examined throughout

the book, in the light of the findings of the review of existing qualitative secondary studies. Before I describe the studies that were identified, in the next chapter I first of all examine the context in which qualitative secondary analysis emerged and became, to borrow Denzin and Lincoln's (1998) phrase, part of the 'landscape' of qualitative research.

## Notes

1. Some of this literature has focused on the potential of re-using statistical data in particular areas of social research, such as health (see Adams et al., 1994; Gleit and Graham, 1989; Gooding, 1988; Herron, 1989; Jacobson et al., 1993; McArt and McDougal, 1985; Woods, 1988), education (see Burstein, 1978; Glass, 1976; Miller, 1982), and criminology (see Riedel, 2000).
2. While this view is generally accepted, Woods (1988) suggests that the initial analysis of, say, clinical data for research purposes is in fact primary analysis. She [p. 17 ↓] therefore retains secondary analysis as a term applying only to data analysed in the course of a prior study.
3. These resources have been described in more detail elsewhere and are simply summarized here (see Dale et al., 1988; Elder et al., 1993; Hakim, 1982; Hyman, 1972).
4. For example, Lobo observes that the following administrative data can be subject to secondary analysis:

Nurses collect or generate an immense amount of data during research and in clinical practice. Many times these data may or may not have been analysed during the initial evaluation of the data. These data may or may not have been organized in a systematic manner when they were initially collected. Clinical data available to nurses for research include nursing notes, vital sign sheets, and input and output sheets. The documents created as a result of nursing research and nursing practice can provide an immense amount of information about diagnosis and treatment of human responses to actual or potential health

problems. These documents can be used as resources and can provide information about what kinds of actions nurses take and the impact of those actions on their individual clients. The research strategy used to analyze material for purposes other than that for which it was originally collected is termed secondary analysis. Although this strategy is usually considered when data collected specifically for research purposes are available, as nurses, we also have access to a large amount of data that are naturally generated as a result of our practice. (1986: 295)

5. The distinction between 'quantitative' and 'qualitative' data, and 'naturalistic' and 'non-naturalistic' data, is not meant to suggest that the re-use of these materials is necessarily segregated in social research. On the contrary, qualitative data may be collected in what are primarily quantitative studies and subject to secondary analysis, as in the case of open-ended responses to survey questionnaires; qualitative data may be subject to more quantitative forms of content analysis. In addition, naturalistic and non-naturalistic data may be combined and jointly analysed in qualitative research studies.

6. The term 'naturalistic' is here used after Plummer (2001) who uses it to distinguish three broad types of life stories: 'naturalistic life stories' which are told as part of everyday life; 'researched and solicited stories' which are mediated by researchers; and 'reflexive and recursive life stories' which are concerned with the telling of the story as well as the story itself.

7. Corti and Thompson (1998) report that interviews from the latter's study, 'Family life and work experience before 1918', archived at the University of Essex since the 1970s, have been seen by over one hundred researchers and students and re-used in numerous books and articles. Atkinson reports that the use of life stories deposited with the Centre for the Study of Lives at the University of Southern Maine in the United States from 1988 was 'slow to develop' (Atkinson, 1998: 4).

8. Denzin (1970) reports that a review of 22 studies that used the life history method between 1920 and 1940 was carried out by Angell (1945). Most of the classic studies using life stories appear to have been based on those collected for primary research purposes rather than those from pre-existing collections. For example, one of the best known studies using a life story is Thomas and Znaniecki's (1958) *The Polish Peasant*

*in Europe and America*. One of its original five [p. 18 ↓ ] volumes (published 1918–20) comprises a 300 page life story of Wladek Wisniewski written shortly before the outbreak of World War I, covering his early life in Poland through to his experiences as a Polish immigrant living in America. His life story was solicited for the study and analysed together with documents in the form of letters, newspaper articles and official public records. Strauss and Glaser (1977) also gathered Mrs Abell's life story for their study: *Anguish: A Case History of a Dying Trajectory*.

9. This is distinct from replication or re-studies in which new data *are* collected in order to re-investigate the questions examined in previous work (see Chapter 3 for more on this topic).

10. For a useful further discussion of related epistemological issues, see Bryman (1988), Hammersley (1997b) and Seale (1999).

11. It may be added that there is a growing literature specifically on techniques for synthesizing the results of quantitative and qualitative research in medicine and, increasingly, in other areas of social research (see Dixon-Woods et al., 2001; Jensen and Allen, 1996; Paterson et al., 2001; Popay et al., 1998).

12. While the possibility of doing meta-analysis using raw data has been recognized for some time (see Woods, 1988: 335), it is still not the norm in meta-research.

13. For an exception, see Herron who, writing on the secondary analysis of quantitative data, has suggested that these data 'may have been gathered earlier and then re-examined by the same researcher...' (1989: 66).

14. While it is beyond the scope of this book to examine secondary studies of quantitative data, it would be interesting to update Hyman's (1972) review of the nature and use of quantitative secondary analysis. In particular, it would be useful to establish the extent to which re-analyses (and re-studies) of research based on quantitative data have been conducted in practice in the social, natural and biomedical sciences.

<http://dx.doi.org/10.4135/9781849209878.n1>