Secondary analysis of qualitative data: an overview
Heaton, Janet

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more Information see:
https://creativecommons.org/licenses/by/4.0

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-191439
Secondary Analysis of Qualitative Data:
An Overview

Janet Heaton *

Abstract: »Sekundäranalyse qualitativer Daten: Ein Überblick«. This paper provides an overview of secondary analysis of qualitative data, based on work examining the nature and use of the methodology in social research. It clarifies what secondary analysis is and how the methodology relates to other similar approaches used in qualitative research. It looks at the development of secondary analysis in qualitative research, and some of the factors that have shaped this. And it examines the ways in which researchers have re-used qualitative data in published studies to date, describing which sources of data have been re-used, by whom and for what purposes. As well as reflecting on the ways in which researchers have so far utilized qualitative data in secondary studies, the paper also looks forward to some of the challenges that lie ahead.

Keywords: secondary analysis; qualitative data; data archiving; re-using qualitative data.

1. Introduction

Since the mid-1990s, there has been growing recognition of and interest in the potential for carrying out secondary analysis of qualitative data. In this article I reflect on how the methodology has been defined, developed and utilized to date, drawing on my work examining the nature and use of qualitative secondary analysis in social research. I begin by describing the main features of the methodology and showing how these distinguish it from other approaches used in qualitative research. Then I briefly outline some of the factors that have facilitated the emergence and ongoing development of secondary analysis in qualitative research since the mid-1990s. In the next part of the paper I describe the ways in which qualitative data have been re-used in published studies. Here I show which sources of data have been re-used by whom, and I describe the different types of secondary analysis represented by existing published work involving the re-use of qualitative data. Finally, I highlight some of the key

* Address all communications to: Janet Heaton, Social Policy Research Unit, University of York, Heslington, York, North Yorkshire, YO10 5DD, United Kingdom; e-mail: jh35@york.ac.uk. URL: <http://www.york.ac.uk/inst/spru/>.
My research on secondary analysis of qualitative data was supported by awards from the Economic and Social Research Council (ESRC) Research Grants Scheme (ref: R000222918) and University of York Innovation and Research Priming Fund (ref: 404). I would like to thank the editors of this volume for their helpful comments on an earlier draft.
issues in qualitative secondary analysis today. Hopefully, the paper provides an insight into the nature and use of secondary analysis of qualitative data to date, and also an idea of the main challenges that lie ahead for future policy and practice concerning the collection, retention, sharing and re-use of qualitative data in social research.

2. What is secondary analysis of qualitative data?

When I began my research on secondary analysis of qualitative data, the first thing I realized was that there was no clear, accepted, definition of the methodology. In the existing literature, secondary analysis of both quantitative and qualitative data was defined in various ways, and different claims were made as to what these methodologies could be used for in social research. I also found some confusion and potential overlap in existing conceptualizations of methodologies for analyzing different types of pre-existing qualitative data, and also of methodologies for analyzing qualitative data as opposed to published findings of research using qualitative methods. These observations led me to try and clarify how secondary analysis related to other methodologies used in qualitative research, and to establish how researchers had adopted and utilized secondary analysis in practice, by reviewing existing publications involving the re-use of qualitative data. The early findings of this work were presented in an issue of Social Research Update (HEATON 1998) and the full findings were published in a final report (HEATON 2000) and formed the basis of a book: Reworking Qualitative Data (HEATON 2004). Below I describe the main features of qualitative secondary analysis that this work has revealed.

2.1 Re-use of pre-existing research data

Secondary analysis involves the re-use of pre-existing qualitative data derived from previous research studies. These data include material such as semi-structured interviews, responses to open-ended questions in questionnaires, field notes and research diaries. In Reworking Qualitative Data, I argued that this focus on non-naturalistic qualitative data distinguishes secondary analysis from documentary analysis (PLUMMER 1983, 2001), which involves working with naturalistic or ‘found’ materials, such as auto-biographies, personal diaries and photographs. However, some types of qualitative material, such as life stories and diaries, may be subject to either secondary analysis or documentary analysis, depending on to what extent the material was solicited and shaped by researchers’ involvement in collecting the material. In revisiting the actual data, secondary analysis is also distinct from meta-analysis (GLASS 1976) and systematic reviews (POPAY, ROGERS & WILLIAMS 1998) of quantitative and qualitative research because these approaches usually involve going back over...
the published findings of previous studies and not revisiting and reworking the data.

2.2 Functions of secondary analysis

In principle, secondary analysis can be used for two main purposes: it can be used to investigate \textit{new or additional} research questions; alternatively, it can be used to \textit{verify} the findings of previous research. Whereas in quantitative research both are accepted functions of secondary analysis, in qualitative research the latter is more controversial, reflecting differences in the epistemological foundations of quantitative and qualitative research, and across different traditions of qualitative research (HEATON 2004). Later, in section 4.2 of this paper, I outline different types of secondary analysis that have been employed in practice, and describe the extent to which each has been used to date.

2.3 Modes of secondary analysis

There are three main modes of secondary analysis. The first is through \textit{formal data sharing}. Here researchers access datasets deposited in public or institutional archives and re-use them in secondary research. In this mode of secondary analysis, researchers are re-using data that were independently collected by others. These data are likely to have been well documented for archiving purposes and to have met the necessary ethical and legal requirements for being shared with other researchers, possibly subject to certain conditions being met. In re-using this source of data, secondary analysts may contact and consult with the primary researchers where they are available, but the latter are not usually involved to the extent that they are named as co-authors of publications arising from the secondary work.

Secondary analysis may also be carried out using datasets obtained through \textit{informal data sharing}. Here there are different possibilities: primary researchers may hand over their data to others, with the former having no part in the actual secondary analysis; or, primary researchers may share their data with others who were not involved in the primary research, and lead or be part of the secondary analysis team; or, two or more primary researchers may get together and pool their own datasets that they collected separately, and work with other independent researchers in carrying out secondary analysis. Thus, whereas in the first option, none of the secondary analysts were involved in the primary research, in the other options, the secondary analysis team includes a mix of researchers who were and were not involved in the original research. Their relationship to the data are therefore different: those that were involved in the primary research bring an inside knowledge of the context in which the primary research was carried out to the secondary analysis.

In the third and final mode of secondary analysis, researchers may re-use their own \textit{self-collected data} in order to investigate new or additional questions
to those explored in the primary research or, alternatively, to verify their previous findings. Here there are two options: researchers can team up and combine their respective datasets (which involves an element of informal data sharing as outlined above, but here only researchers who were involved in one or more of the primary studies are involved in the secondary analysis); or, a sole researcher or team of primary researchers can re-use data that they themselves previously collected for secondary research purposes (and again without involving any other independent parties).

Later, in section 4.1, I show which sources of data have been re-used in existing studies, but first I look at which modes of secondary analysis have been promoted and developed internationally to date.

3. Development of the methodology

While secondary analysis is a well known and well established methodology for re-using quantitative data in social research (DALE, ARBER & PROCTER 1988; HAKIM 1982), the development and utilization of secondary analysis of qualitative data has been slower and more controversial. The potential of reusing qualitative research data was recognised in the 1960s (GLASER 1962, 1963), but it was not until the mid-1990s that the first article dedicated to the secondary analysis of qualitative data was published by Sally THORNE (1994) and that studies based on the re-use of qualitative research data began to be published in North America and, to a lesser extent, Europe and Australia.

Interest in the possibilities of secondary analysis of qualitative data has grown in the past decade, kindled by various developments. In USA and Europe, data archives have increasingly begun to accept non-naturalistic qualitative datasets from research studies and to make them available for further use in secondary research. Previously data archives had concentrated on collecting and sharing statistical datasets. Where archives were interested in qualitative material, this was mainly confined to oral histories and cultural artefacts, such as diaries and other naturalistic documents (for example, The Institut für Geschichte und Biographie¹ and The Verein zur Förderung von Forschungen zur politischen Sozialisation und Partizipation (POSOPA e.V.)² in Germany; and the Tom Harrison Mass Observation archive³ in UK), or involved the collection of qualitative material that was then coded for statistical analysis (for example, the Human Relations Area Files⁴ project in USA).

¹ Available at: <http://www.fernuni-hagen.de/INST_GESCHUBIOG/welcome.shtml>.
² Available at: <http://www.posopa.de/>.
³ Available at: <http://www.massobs.org.uk/index.htm>.
⁴ Available at: <http://www.yale.edu/hraf/>.
A notable landmark in the history of data archiving was the establishment of Qualidata, the world’s first organisation for promoting archiving and re-use of qualitative data, in 1994 at the University of Essex, UK. It was set up by the Economic and Social Research Council (ESRC) and now forms part of the Economic and Social Data Service (ESDS). Since then, ESDS Qualidata and ESRC have been proactive in promoting and facilitating archiving and re-use of qualitative data derived from classic studies and ongoing research in UK.

Internationally, policies promoting data retention and sharing in social research have been gradually extended to include qualitative datasets. In the UK, for example, ESRC has introduced policies that require researchers to deposit qualitative datasets from primary research it has funded, and which encourage new applicant’s to draw wherever possible on secondary sources to answer proposed questions in their bids. More generally, other funders of research in UK have updated policies regarding data retention in order to ensure that datasets are kept for a reasonable period beyond publication to at least enable them to be available for verification of results if required.

At the same time, advances in computing have made preparing qualitative data for archiving, and improving access to datasets for secondary research purposes by independent researchers, more feasible. Further advances in digitization of data are likely to improve opportunities for preserving qualitative data, and enabling easier and wider access to datasets on-line in the future.

While official policies have increasingly promoted qualitative data archiving and formal sharing of these data, in a parallel development a growing number of researchers from North America working in the area of health and social care research have begun to publish secondary studies drawing on qualitative data obtained from non-archived sources, namely informally shared data and self-collected datasets. These researchers have pioneered approaches to reusing qualitative data which have in turn informed subsequent studies by the same and other researchers.


---

Available at: <http://www.esds.ac.uk/qualidata/about/introduction.asp>.
in UK around the development of ESDS Qualidata and ESRC policies, concerning the place of secondary analysis in the landscape of qualitative research – which some have questioned (see BISHOP 2005; MAUTHNER, PARRY & BACKETT-MILBURN 1998; PARRY & MAUTHNER 2004, 2005). I shall return to some of the key issues raised later in this article, following an outline of the ways in which researchers have re-used qualitative data in published studies to date.

4. Existing approaches to re-using qualitative data

As I mentioned earlier, when I began my research on secondary analysis of qualitative data, there was no agreed definition of the methodology and only a few articles had been published then exploring the possibilities of re-using qualitative research data. I was keen to find out which qualitative research data had been re-used, by whom and for what purposes in existing studies, and so began to search for and review such work. My review concentrated on the international health and social care literature, where my early searching had revealed that most secondary studies had been published, and subsequently I extended my work to include studies in other areas, such as criminology and education.

4.1 Which sources of data have been re-used and by whom?

By the time Reworking Qualitative Data was published, I had identified 65 publications based on studies involving the re-use of qualitative data. These included 41 that were clearly defined by the authors as involving secondary analysis. The remainder were either described in other terms, for example, as ‘post hoc analysis’, or the approach was not defined at all but displayed the characteristics of secondary analysis.

The majority of studies originally reviewed (n=51, 78%), were carried out by researchers from North America; 12 were by authors from UK, one from Sweden, and one was a pan-Canadian/Swedish study. Since the formal review was carried out, further secondary studies have also been published by researchers from other European countries, such as France (DARGENTAS & LE ROUX 2005), Finland (HÄGGMAN-LAITILA 2005), Germany (MEDIJEVIĆ & WITZEL in this volume), and from Australia (IRURITA & WILLIAMS 2001; KELDER 2005; KRALIK, BROWN & KOCH 2001; KRALIK, KOCH & TELFORD 2001).

In most studies (n=56, 86%), one or more authors had also been involved in the primary research from which datasets were derived. These secondary studies were carried out using data obtained through informal data sharing or using self-collected datasets from primary research that the authors had themselves previously carried out as sole researchers or as part of a research team. Exam-

In only nine (14%) studies did the authors appear to have had no involvement with the original research and to have carried out the secondary work independently of the primary researchers; just two of these studies involved use of publicly archived datasets (BEVAN 2000; BLOOR 2000). Further studies have since been identified in other areas involving the re-use of such datasets, notably ones by FIELDING & FIELDING (in this volume) and SAVAGE (2005a & 2005b).

4.2 Types of secondary analysis

Whereas previous work had pointed to potential ways in which qualitative data could be re-used (for example, see HINDS, VOGEL & CLARKE-STEFFEN 1997; THORNE 1994), the review provided an empirically based classification of existing approaches to secondary analysis of qualitative data. Looking at the characteristics of the studies reviewed, I initially identified six types of secondary analysis (HEATON 2000) and later collapsed these to five (HEATON 2004). My typology included some approaches that were similar to those previously identified, and some that were new.

The first three types of secondary analysis identified vary according to the degree to which the aims of the primary and secondary work converge or diverge. In supplementary analysis, a more in-depth analysis of an emergent issue or aspect of the data, that was not addressed or was only partially addressed in the primary study, is undertaken. This was the most common form of qualitative secondary analysis, with 39 such studies identified. It differs from supra analysis, where the aims and focus of the secondary study transcend those of the original research; 14 such studies were identified. Only one example of the third type of secondary analysis, re-analysis, where data are re-examined in order to confirm and validate findings of a primary study, was found. The last two types of secondary analysis intersect with the above and are distinguished by the number and nature of datasets involved. Thus, in amplified analysis, two or more existing qualitative datasets may be compared or combined for purposes of secondary analysis; 17 such studies were found, 14 of which made use of two datasets. And in assorted analysis, re-use of existing qualitative data is carried out alongside the collection and analysis of primary qualitative data for the same study, or a mix of naturalistic and non-naturalistic data are examined; five examples of this type were identified.
5. Issues in qualitative secondary analysis

In the UK, two efforts to determine researchers’ attitudes to archiving and re-using qualitative data provided conflicting findings. A survey, carried out by Qualidata in 1999, found that a large majority of respondents wanted access to archived qualitative datasets (CORTI & THOMPSON 2004). However, a consultation on ESRC Data Policy and Archiving undertaken two years later found mixed support for, and considerable concerns about, the archiving and re-use of qualitative data among the research community (BODDY 2001). Three main areas of debate have dominated the literature.

First, some commentators, including researchers and archivists, have debated whether the re-use of qualitative data is in keeping with some of the fundamental principles of qualitative research (BISHOP 2005; CORTI 2003, 2006; FIELDING 2004; HAMMERSLEY 1997; HEATON 1998/2003, 2000, 2004, 2008; MAUTHNER, PARRY & BACKETT-MILBURN 1998; PARRY & MAUTHNER 2004, 2005; VAN DEN BERG in this volume). Key issues include the ‘problem of data fit’ – or whether data collected for one (primary) purpose can be re-used for another (secondary) purpose. In qualitative research, the relatively flexible nature of research designs, and semi-structured nature of data collected, can result in datasets with variable depth and breadth of coverage of topics, hence the need to ensure that data can be used to meet the new aims of a secondary study. Another major issue is the ‘problem of not having “been there”’ – where analysts try to interpret data that were collected by other researchers. This problem is not necessarily confined to the interpretation of qualitative data from archives, but includes secondary analysis of informally shared data as well as primary analysis of data collected by teams where different researchers carry out interviews and jointly analyse them. And a third key issue is the ‘problem of verification’ – or whether the results of qualitative research can or should be verified in the same ways as studies using statistical methods. Here tensions in the epistemological foundations of quantitative and qualitative research, and across different traditions of qualitative inquiry, are most apparent. Methods of verification derived from positivist-based approaches, which underpin the data sharing imperative in quantitative research, are anathema to many qualitative researchers, for whom alternative methods have been developed to help establish the ‘trustworthiness’ and authenticity of their work (for example, see LINCOLN & GUBA 1985).

Second, a number of ethical and legal issues have been raised, such as whether and how best to obtain informed consent from research participants for retaining data, sharing data with others, and re-using it for purposes other than those for which it was originally collected; how to maintain confidentiality of data and, related to this, how to anonymize data without rendering it bare; and how to act in accordance with laws on copyright and data protection (CORTI,
DAY & BACKHOUSE 2000; PARRY & MAUTHNER 2004; THORNE 1998). In UK, guidelines on these issues, in relation to the creation and archiving of qualitative datasets, have been developed by ESDS Qualidata and posted on its website; ESRC also commissioned a review of the legal issues involved (ALLEN & OVERY 1998).

Finally, I have argued that existing policy and guidelines concerning the retention and re-use of qualitative data have to date focused almost exclusively on publicly archived data and needs to be more inclusive of other sources of secondary qualitative data too – that is, informal data sharing and self-collected datasets – which have been shown to be the main sources of data being used in published studies to date and potentially a useful resource in the future (HEATON 2008).

6. Conclusion

Since the mid-1990s, interest in the potential of secondary analysis of qualitative data has grown internationally; more researchers have published secondary studies involving the re-use of qualitative data; and opportunities to access and re-use qualitative data from data archives, informal data sharing and self-collections, have greatly increased.

The findings of my review of existing qualitative secondary studies have shown that researchers have so far tended to re-use their own data, either alone or by informally sharing their work with others, rather than drawing on independently collected and formally archived datasets. While the review provided evidence of which sources of data researchers have drawn on to date in their published work, it did not establish whether this was because informal and private sources were preferred over archived sources or if, say, appropriate archived datasets were not yet available for researchers to use. However, it did reveal the potential of re-using non-archived as well as archived sources of qualitative data in the future.

Official policies in UK and elsewhere have tended to promote secondary analysis of qualitative data through developing resources for archiving and formal data sharing, while in practice researchers have so far mainly re-used data obtained through informal data sharing or from self-collections. In developing future policy concerning the re-use of qualitative data, and in debating the whys and wherefores of secondary analysis of qualitative data, it is important that attention is paid to all the different sources that researchers may use – archives, informal data sharing networks and self-collected datasets – as the issues involved are different, depending on which sources are used and researchers’ affiliation to these data.
References


