SECONDARY ANALYSIS OF QUALITATIVE DATA: A REVIEW OF THE LITERATURE

REF: R000222918

Janet Heaton
Research Fellow
Social Policy Research Unit
University of York

August 2000
SECONDARY ANALYSIS OF QUALITATIVE DATA: A REVIEW OF THE LITERATURE

ESRC 1752 JH 8.00

Janet Heaton
Research Fellow
Social Policy Research Unit
University of York

REF: R000222918
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SUMMARY OF RESEARCH RESULTS</td>
<td>1</td>
</tr>
<tr>
<td>2. FULL REPORT OF RESEARCH ACTIVITIES AND RESULTS</td>
<td>4</td>
</tr>
<tr>
<td>Background</td>
<td>4</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
</tr>
<tr>
<td>Methods</td>
<td>5</td>
</tr>
<tr>
<td>Results</td>
<td>6</td>
</tr>
<tr>
<td>Activities</td>
<td>19</td>
</tr>
<tr>
<td>Outputs</td>
<td>19</td>
</tr>
<tr>
<td>Impacts</td>
<td>20</td>
</tr>
<tr>
<td>Future research priorities</td>
<td>20</td>
</tr>
<tr>
<td>3. REFERENCES</td>
<td>21</td>
</tr>
<tr>
<td>4. APPENDIX: ‘SECONDARY ANALYSES OF QUALITATIVE DATA: AN ANNOTATED BIBLIOGRAPHY’</td>
<td>26</td>
</tr>
</tbody>
</table>
SUMMARY OF RESEARCH RESULTS

Secondary analysis (SA) is a term used to describe various analytical practices which use pre-existing data to investigate new research questions or to re-examine primary study questions for purposes of corroboration. While the SA of quantitative data is an established and well documented methodology, re-use of qualitative data remains under-developed.

This study had three main aims:

• to describe the nature and use of SA of qualitative data sets in the international health and social care literature;
• to critically appraise these studies, examining the epistemological, methodological and ethical issues arising from the use of SA in this area of research;
• to consider the overall quality and impact of the studies and implications for future development of the methodology.

The nature and use of qualitative secondary analysis

Fifty-five studies were identified which were self-defined as SA (N=36) or which fitted the above definition (N=19). The vast majority of studies (N=49) were published in the 1990s and were by authors based in North America (N=47). Only seven studies originated from the UK. A large proportion of the studies (N=48) were authored by at least one person who was involved with the primary research from which the data were derived. Thus, researchers were tending to re-use their own data, sometimes sharing it informally with others, rather than drawing on independently collected and archived data sets.

Appraisal of these studies has revealed six types of SA being undertaken:

Supra analysis: transcends the focus of the primary study from which the data were derived, examining new empirical, theoretical or methodological questions.

Amplified analysis: combines data from two or more primary studies for purposes of comparison or enlarging sample.

Supplementary analysis: a more in-depth investigation of an emergent issue or aspect of the data which was not addressed in the primary study.

Complementary analysis: the SA is supported by additional primary research or, alternatively, a primary study which includes an element of SA.
Alternative analysis: data are re-analysed using new methods and/or perspectives for purposes of corroboration based on the principle of triangulation.

Repeat analysis: data are re-analysed using a similar analytical framework in order to verify the findings of the primary research.

Issues arising from the appraisal of studies

Qualitative SA was found to differ from its quantitative counterpart in that it places emphasis on the re-use of data sets to investigate new research questions (rather than the re-investigation of primary study questions for purposes of corroboration). This may reflect the different epistemological foundations of research using qualitative and quantitative methods; thus, the latter methods are mainly used in the more positivistic social, natural and biomedical sciences where greater emphasis is placed on the need to corroborate findings through re-analysis or re-studies.

The finding that qualitative SA was generally conducted by researchers who were involved in the primary studies from which the data were derived not only further distinguishes the SA of qualitative data from that of quantitative data but also raises issues about the relationship of the analyst to the data. Is the analyst who has direct knowledge and/or experience of the context of the data collection in a better position to conduct a SA? While this may be an advantage, levels of involvement in primary research do vary. In addition, it may be possible for secondary analysts using archived data to consult with the primary researchers. Moreover, all analysts, whether or not they were ‘there’ at data collection, produce (primary and secondary) analyses which are socially contingent.

Poor reporting of the conduct of the studies meant that it was not possible to assess their overall quality. Many failed to report on key issues, including the selection of the data set(s) and sub-sample(s), whether informed consent was sought from informants, how the ‘fit’ of the data and the focus of the SA was ascertained, how the rigour of the analysis was established, and the possible limitations of the methodology. Where there was evidence that researchers had made efforts to establish the rigour of their analyses, it was not clear from the reports how the techniques employed, most of which are used in primary research, were transferred and applied in the context of SA. However, there was evidence of methods, such as grounded theory, being adapted specifically for use in this context.

Only two studies reported seeking informed consent from the informants whose data was used for the SA. As the majority of studies failed to discuss the topic, it is not clear if consent
was simply presumed or if it was sought but not reported. This finding raises the issue of whether informed consent should be obtained in all secondary analyses and, if so, how. A related concern is whether the sharing of data violates prior agreements concerning the confidentiality of the information. The evidence from this review suggests that it may be inappropriate to generalise about the need to obtain informed consent for secondary analyses, as this is likely to vary according to the characteristics of the secondary study.

**Implications for future development of the methodology**
The following implications for the future development of the methodology were identified:

- There is a need to better define the use of SA in reports and in electronic databases.
- There is a need more fully to document the ways in which SA was accomplished in reports of secondary studies.
- The possibility of qualitative researchers extending their primary work through the SA of *auto-data* or through *informal data sharing* should be acknowledged and facilitated.
- Social science professional guidelines on the procurement of informed consent and preservation of confidentiality need to be revised so that they consider these issues more specifically in the context of secondary research.
- Further research is required:
  - on the use of SA of qualitative data in other subject areas;
  - on researchers’ attitudes toward the use of archived qualitative data;
  - to establish what constitutes a worthy qualitative data set for archiving and re-use;
  - on ‘how to do’ SA of qualitative data;
  - on the public understanding of social science research and, in particular, lay views on the re-use of data. This includes the circumstances in which informed consent should be sought for secondary research, as well as the meaning of confidentiality and appropriateness of related techniques for anonymising data.
BACKGROUND
Secondary analysis (SA) is a term used to describe various analytical practices which make use of pre-existing data either to investigate new research questions or to re-examine primary study questions for purposes of corroboration.\(^{(1-13)}\) Data used in this context are thereby transformed into ‘secondary data’. This data, which can be quantitative or qualitative or a mixture of the two, may have been collected in the context of a research study (hereafter referred to as ‘data sets’) or for other purposes (referred to as ‘naturalistic’ data, such as diaries and administrative records). Although meta-analysis and literature reviews involve the use of raw and/or published data from existing studies, these methodologies can be distinguished from SA in that they are concerned with synthesising research findings, rather than examining new research questions or verifying results from individual studies.\(^{(4, 12)}\)

This study is concerned with the SA of qualitative data sets. Unlike the SA of quantitative data, which is an established and well documented methodology,\(^{(1, 3, 6, 7)}\) the re-use of qualitative data sets remains under-developed. Although the potential for re-using this type of data was recognised as long ago as 1962,\(^{(14)}\) SA has been mainly confined to statistical data from large scale social surveys and other sources. However, recent developments in the UK have both galvanised interest in the SA of qualitative data sets and made the methodology a more realistic proposition. In particular, the establishment of Qualidata by the Economic and Social Research Council (ESRC) in 1994 and the adoption of policies promoting qualitative data archiving and re-use by the ESRC and other research funding organisations has created the conditions of possibility for increased sharing of qualitative data. There is therefore a need to examine existing efforts to pioneer the SA of qualitative data sets and to consider how the methodology might best be developed in the future. This review builds on the emerging literature on this topic, which has begun to outline the rudiments of the methodology and its potential uses,\(^{(10-12)}\) as well as the different forms it can take.\(^{(4, 5)}\)

OBJECTIVES
The review had three main objectives:
• to describe the nature and use of SA of qualitative data sets in the international health and social care literature;
• to critically appraise these studies, in particular examining the epistemological, methodological and ethical issues arising from the use of SA in this area of research;
• to consider the overall quality and impact of the studies and the implications for future development of the methodology.

The focus on the international health and social care literature was chosen for two main reasons. First, it was already known that this was an area where SA of qualitative data had been developed.\(^4\) Secondly, the range of methods utilised, and the sensitive nature of some of the topics covered in this area, made it ideal for exploring the theoretical, technical and ethical issues arising from the re-use of qualitative data sets. All the objectives were addressed, although the poor reporting of the methods used in the studies limited the assessment of the quality of this work and related issues, such as the possible influence of newly developed computer software for analysing qualitative data. The recency of many of the studies also restricted the extent to which their impact could be considered.

**METHODS**

A number of resources were searched (electronically and manually) to identify secondary analyses of qualitative data: electronic databases (including BIDS Social Science Citation Index, CINAHL, EMBASE, Medline, and Sociological Abstracts); e-journals; electronic search engines; publishers’ electronic bibliographies; key journals; study references; and key author citations. Appeals for information were also issued via the British Sociological Association *Network* magazine, conference flyers, electronic discussion lists (QUALRS-L and ARCHIVE-QUALITATIVE-DATA), and personal contacts.

As the electronic databases did not keyword ‘secondary analysis of qualitative data’ a number of synonymous terms, such as ‘secondary interview data’ and ‘secondary-analysis with qualitative-studies’, were derived from the studies initially identified. These were used systematically to search study titles and abstracts for examples of relevant studies. Abstracts identified by these means were read and the studies obtained if they:

- employed SA of qualitative data sets (as defined above, p.3);
- were in the area of health and social care;
- were published in English.

Once obtained, the studies were checked for relevance and appraised. In some cases authors were approached to clarify whether or not they defined the study as an example of SA. Details of each of the studies were entered into a spreadsheet (QuattroPro). An annotated bibliography was also compiled using a reference management software package (EndNote) (see appendix).
While the review was wide-ranging, it cannot claim to be exhaustive, mainly because of problems with the definition of SA. As we will see, the review includes several studies which, although not self-defined as SA, nevertheless exemplify characteristics of the methodology. This finding confirms Thorne’s\(^\text{12}\) observation that secondary analyses are not always readily defined as such.

**RESULTS**

The results of the review are presented under three main headings corresponding to the objectives of the study.

**Extent and nature of use of secondary analysis**

A total of 55 studies which were self-defined as SA (N=36) or which fitted the definition (N=19) were identified by the review. Other terms authors employed to describe the latter studies included: ‘post hoc analysis’, ‘reanalysis’ and ‘retrospective latent content analysis’.

The vast majority of studies (N=49) were published in the 1990s and were by authors based in North America (N=47). Only seven studies originated from the UK. While allowing that these findings may reflect the geographical coverage of some of the resources searched and the limitations of the search strategies adopted (which excluded studies not published in English), they nevertheless suggest that the methodology is more developed in North America than elsewhere.

Studies were classified into one of three groups depending on the relationship of the author(s) to the data. The largest group of studies (N=32) were by ‘non-independent’ authors - analysts who had been involved in the primary research and were therefore re-using their own previously collected data (hereafter referred to as ‘auto-data’). In the second group of 16 ‘semi-independent’ studies one or more - but not all - of the authors had been involved in the primary study. Thus, 48 out of the 55 studies were authored by at least one person who was involved with the primary research. Of the remaining studies, six were classified as ‘independent’ secondary analyses as the author(s) had not been involved with the primary research, and one study used both auto - and independent data sets.

The majority of studies (N=40) were secondary analyses of a single data set. Twelve studies drew on two data sets and three utilised more than two data sets in total. While secondary face-to-face interview data were utilised in 46 studies, qualitative data derived from observational work, focus groups, surveys, telephone interviews, vignettes, documents, autobiographies and published ethnographies were also drawn upon. In half the secondary analyses a subset of the data from the primary study was selected for re-use; the other half
utilised the full data set. A small number of studies (N=5) also involved the collection of additional primary or secondary (naturalistic) data.
Previous overviews of the methodology have drawn attention to various possible types of SA of qualitative data.\(^4\)\(^,\)\(^5\)\(^,\)\(^11\) Building on this work, the studies identified in this first extensive empirical review of secondary analyses were examined for variations in their approach. The studies were found to vary on five main dimensions:

- **function** - whether or not the SA was designed to investigate new research questions or to re-investigate a primary study question for purposes of corroboration;
- **focus** - how the question(s) addressed by studies investigating new research questions differed from those addressed in the primary research;
- **source(s) of data** - whether one or more secondary data sets were used and whether any primary research was also conducted alongside the SA;
- **methods/perspectives** - whether these were similar to or different from the approach employed in the primary research;
- **origins of secondary data** - the extent to which studies relied on auto-data and/or data from other sources.

Based on this information, six varieties of SA were discerned: supra analysis, amplified analysis, supplementary analysis, complementary analysis, alternative analysis and repeat analysis. While some of these types of SA are similar to those previously identified, others are new. Hence, this typology refines and extends previous models of classification, providing knowledge of the range of approaches to the SA of qualitative data in the area of health and social care research.

Following an adumbration of the typology in Table 1, each of the types are fully described below together with some illustrative examples of relevant studies.

*Table 1: Typology of secondary analysis of qualitative data*

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supra analysis</td>
<td>Transcends the focus of the primary study from which the data were derived, examining new empirical, theoretical or methodological questions.</td>
</tr>
<tr>
<td>Amplified analysis</td>
<td>Combines data from two or more studies for purposes of comparison or enlarging sample.</td>
</tr>
<tr>
<td>Supplementary analysis</td>
<td>A more in-depth investigation of an emergent issue or aspect of the data which was not addressed in the primary study.</td>
</tr>
<tr>
<td>Complementary analysis</td>
<td>The SA is supported by additional primary research; or a primary study which includes an element of SA.</td>
</tr>
<tr>
<td>Alternative analysis</td>
<td>Data is re-analysed using new methods and/or perspectives for purposes of corroboration based on the principle of triangulation.</td>
</tr>
<tr>
<td>Repeat analysis</td>
<td>Data is re-analysed using a similar analytical framework in order to check the findings of the primary research.</td>
</tr>
</tbody>
</table>
It is important to note that these are ideal types and hence individual studies may not always exhibit all the characteristics of a given type and, indeed, may share attributes of different types. Accordingly, studies have been classified as exemplifying a particular type (or types) of SA on the basis of their principal characteristics. A summary of the number and key characteristics of the different varieties of SA are set out in tabular form in Table 2 at the end of this section.

Supra analysis

‘Supra analyses’ transcend or go beyond the terms of the primary or ‘parent’ study from which the data were derived. Such studies involve the investigation of new empirical, theoretical or methodological questions. An equivalent approach was previously highlighted (but unnamed) by Heaton\(^4\) and shares some characteristics with a variety of SA called ‘armchair induction’ identified by Thorne,\(^{11}\) although supra analysis may be conducted by the same researchers who carried out the primary research and is not necessarily restricted to theoreticians.

Eight out of the 55 studies developed analyses which transcended the terms of the primary work, focussing on new aspects of the data and often employing a new theoretical perspective.\(^{15-22}\) Although not self-defined as SA, a study by Bloor and McIntosh\(^{15}\) usefully exemplifies the characteristics of supra analysis (as well as amplified analysis - see below). They analysed data from two separate studies of health visiting and therapeutic communities. The focus of their analysis was clearly not planned at the outset of the study:

‘Of course, the health visiting and therapeutic communities studies were not undertaken in order to compare professional-client relationships in different types of service-provision, nor were they undertaken to elaborate a Foucauldian approach to client resistance. The possibility of writing this chapter only occurred to us post hoc, when we realised that we both had data bearing upon issues of power and contest which showed both similarities and dissimilarities in techniques of client resistance.’ (p.161).

In focussing on forms of surveillance in professional-client relationships and associated strategies of resistance, the authors go beyond the terms of the original studies and develop a new, free-standing analysis consistent with supra analysis as it is defined here.

Amplified analysis

This approach draws on two or more existing research data sets in order to examine common and/or divergent themes across them. It can involve the comparison of different study populations or the pooling of data on a similar population. This definition incorporates variations of this approach previously described as ‘amplified sampling’,\(^{11}\) ‘aggregated
analysis’, \(^{(23)}\) and ‘pooled case comparison’, \(^{(24)}\) which all involve the SA of multiple pre-existing data sets.

Amplified analysis was the second most common type of SA identified by the review. Seventeen examples were found. \(^{(15, 16, 25-39)}\) Of these, 12 made use of two data sets. All the amplified analyses were by authors who had been involved with the primary studies, although additional independent researchers also collaborated with some of these secondary analyses.

Two main groups of amplified analysis were discerned. In the first, researchers who separately conducted independent primary studies later performed a SA of the data sets to further explore common issues across the study populations. For example, Yamashito and Forsyth \(^{(39)}\) learned, through meeting at a conference, that they had each conducted studies of families’ reactions to a relatives’ mental illness in Canada and the USA respectively. This meeting led to a SA in which they draw on ‘aggregated analysis’ \(^{(23)}\) to examine similarities and differences in the two data sets.

In the other group of amplified analyses, researchers made use of two or more data sets from their own oeuvre. This approach is illustrated by a series of studies by Sally Thorne and Carole Robinson. They combined the data from their MA projects and published two secondary studies on health care relationships. \(^{(33, 38)}\) Building on this work, they subsequently obtained funding to do a primary study in 1989. \(^{(40)}\) Individually the authors then produced four secondary analyses of these data sets (co-classified here as examples of amplified/supplementary analyses). Thus, Thorne used the data to examine mothers’ experiences of chronic illness and the meaning of non-compliance for those with chronic illness, as well as completing a doctoral thesis involving the analysis of secondary and primary sources. \(^{(35-37)}\) Robinson later drew on part of the same material to examine the normalisation of life by chronically ill members and their families. \(^{(32)}\) Although this last study is not self-defined as SA, it is set in the context of the aforementioned work and the theme it explores was analysed separately and retrospectively. These examples of the re-analysis of data from one’s own oeuvre reflect an incremental approach to research whereby individual studies evolve progressively, rather than forming a disconnected series of works.

**Supplementary analysis**

‘Supplementary analysis’ involves a more in-depth focus on an emergent issue or aspect of the data which was not addressed (or was only partially addressed) by the primary analysis. Thus, the focus of the SA may shift to a particular theme, or to issues pertaining to a sub-sample of the primary study population. Supplementary analysis is in these respects related to Thorne’s ‘retrospective interpretation’ and ‘analytic expansion’ categories (the former referring specifically to researchers’ re-use of their own data). \(^{(12)}\) Other examples of this variety of SA have been previously described by Heaton \(^{(4)}\) and Hinds et al. \(^{(5)}\)
As the foci of supplementary analyses are compatible with that of the primary study, the two may be difficult to separate, particularly as researchers are increasingly selectively reporting the results of primary research in multiple publications rather than in a single, overarching publication. However, while selective reporting may focus on a particular aspect of the primary study, it is still produced in the context of the ongoing primary research. By contrast, supplementary analysis exceeds the analytical remit of the primary study, examining in more depth a theme or sub set of the data which has emerged as a post hoc matter of interest and subsequent focus of inquiry.

This approach was found to be by far the most common form of SA of qualitative data, exemplified by 36 of the 55 studies identified. In the majority of the supplementary analyses (N=22) a subset of the data from the primary study was examined. For example, in one of two secondary analyses of data from their primary study of women cocaine users, Kearney et al. focussed in more depth on the mothers in this sample. The remainder of the supplementary analyses (N=14) involved the re-examination of an entire data set. For instance, following on from one of the authors’ exploratory, descriptive study of pain management at home by cancer patients, Vallerand and Ferrell’s secondary study on the concept of control involved the use of the full data set.

**Complementary analysis**

This form of SA is combined with primary data collection and analysis. Such a study may rely mainly on secondary data, with primary research conducted in support of the former. Equivalent examples of this approach were first identified by Hinds et al. Alternatively, the analysis of secondary data may form a relatively minor part of a primary study. This use of SA has been defined as ‘cross-validation’ by Thorne. Complementary analysis can therefore be used to provide additional comparative or collateral evidence using different sources of data, or to verify the results of the main analysis.

Four out of the five examples of complementary analyses identified were basically secondary studies which involved some additional primary data collection and analysis. Two of these studies - neither self-defined as secondary analyses - were by Hutchinson and made use of data from a study of unprofessional behaviour among hospital-based nurses. In the first ‘descriptive’ study, the complementary/supplementary analysis focussed on nurses’ self-care strategies. A sub-set of the data was re-used and an additional 20 interviews were conducted with nurses to augment the data set and to check initial observations. In the second ‘grounded theory’ study, the complementary/supplementary analysis focussed on how
nurses bend the rules for the sake of the patient. Again a sub-set of the data was re-used and an additional 21 interviews were conducted in order to ‘verify’ the initial observations.
Only one example of a primary study incorporating analysis of secondary data sources was found. Sandelowski and Jones undertook a primary study of couples’ experiences of prenatal diagnosis and, in particular, their responses to the acquisition of foreknowledge of fetal impairment. Additional information from research on parenting medically fragile infants by another investigative team was also used ‘for purposes of further comparison and validation’ (p.85).

**Alternative analysis**
Unlike the preceding four types of SA which are all concerned with the investigation of new research questions, ‘alternative analysis’ maintains a focus on the question(s) posed in the primary study but re-investigates these using new methods and/or analytical frameworks. This approach may therefore produce alternative findings or theoretical insights, or it may confirm and validate the original findings through triangulation. Alternative analysis has not been previously identified in overviews of the SA of qualitative data.

Only one example of this form of SA was found. Re-using data from research on overeating among women who weight-cycle, the study aimed:

‘to analyse further subjects’ interview responses to ensure that no important information was omitted in the primary analysis, which used a reversal theory coding system only. The secondary analysis provided a validity check for the primary coding results and an accuracy check for complete interpretation. Using methodological triangulation... two methods of viewing the same empirical content, two coders performed a content analysis of the interview data with no consideration for reversal theory’ (p.71).

Thus, in this case, the aims of the primary and secondary study were compatible, but the methods of analysis were different.

**Repeat analysis**
Like alternative analysis, ‘repeat analysis’ maintains a focus on the primary study question(s) but instead uses a similar analytical approach with the intention of corroborating the findings of the original research. Repeat analysis may in principle be conducted either by independent analysts or by the original researchers in order to verify their findings. This approach epitomises the spirit of open scientific enquiry which has underpinned the development of the data sharing movement in the USA.

While this type of SA is generally recognised in theory, no examples of it being applied in practice to qualitative data were identified by this review. However, an example of a repeat analysis of this type of data from another area is Lewis’s re-analysis and re-study of Redfield’s research on life in a Mexican village.
Table 2: Principal characteristics of types of secondary analysis

<table>
<thead>
<tr>
<th>TYPE &amp; No.*</th>
<th>FUNCTION</th>
<th>FOCUS</th>
<th>SOURCE(S) OF DATA</th>
<th>METHODS/PERSPECTIVES</th>
<th>ORIGINS OF SECONDARY DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPRA ANALYSIS (N=8)</td>
<td>To investigate a new research question</td>
<td>Transcends primary study (PS)</td>
<td>Secondary data set(s) or mixed secondary and primary data</td>
<td>Discontinuous or continuous with PS</td>
<td>Independent data or auto-data</td>
</tr>
<tr>
<td>AMPLIFIED ANALYSIS (N=17)</td>
<td>To investigate a new research question</td>
<td>Transcends PS or in-depth analysis of an aspect of the PSs</td>
<td>Multiple secondary data sets</td>
<td>Continuous or discontinuous with PS</td>
<td>Independent data and/or auto-data</td>
</tr>
<tr>
<td>SUPPLEMENTARY ANALYSIS (N=36)</td>
<td>To investigate a new research question</td>
<td>In-depth analysis of an aspect of the PS</td>
<td>Secondary data set(s) or mixed secondary and primary data</td>
<td>Continuous or discontinuous with PS</td>
<td>Auto-data or independent data</td>
</tr>
<tr>
<td>COMPLEMENTARY ANALYSIS (N=5)</td>
<td>To investigate a new research question or validate an ongoing primary study (PS)</td>
<td>In-depth analysis of an aspect of the PS or transcends PS</td>
<td>Mixed secondary and primary data sets</td>
<td>Continuous or discontinuous with PS</td>
<td>Independent or auto-data</td>
</tr>
<tr>
<td>ALTERNATIVE ANALYSIS (N=1)</td>
<td>To re-investigate (PS) question</td>
<td>Same focus</td>
<td>Secondary data set</td>
<td>Discontinuous with PS</td>
<td>Independent or auto-data</td>
</tr>
<tr>
<td>REPEAT ANALYSIS (N=0)</td>
<td>To re-investigate PS question</td>
<td>Same focus</td>
<td>Secondary data set; may be part of a wider primary re-study</td>
<td>Continuous with PS</td>
<td>Independent or auto-data</td>
</tr>
</tbody>
</table>

*The numbers add up to more than 55 because some studies shared the principal characteristics of more than one type.
Issues arising from the appraisal of studies

Various epistemological, methodological and ethical issues were raised by the critical appraisal of the studies. These will be discussed under separate headings, although some issues are clearly inter-related.

Epistemological issues

To date, the concept of SA has mainly been defined in relation to the re-use of quantitative data. Most commentators agree that it involves the use of pre-existing data either to investigate new research questions or to re-investigate primary study questions using new methods or alternative perspectives. However, as we have seen, the secondary analyses identified by this review have generally been concerned with the former. This may reflect the different epistemological foundations of research using qualitative and quantitative-based methods; thus, the latter methods are mainly used in the more positivistic social, natural and biomedical sciences where greater emphasis is placed on the need to corroborate findings or replicate studies.

SA is also generally associated with the use of statistical data sets which were originally collected by others\(^1,3,7,9\) although, as Hyman\(^6\) observes, in the USA researchers did re-use their own survey data before data archives were established. While no one since Hyman has considered this possibility in relation to quantitative data, in the emerging literature on the SA of qualitative data it has been recognised that the methodology can in principle involve auto-data as well as independently collected data sets.\(^4,10-12\)

As we have seen, auto-data was used in 48 of the 55 studies identified by the review. In re-using their own data, researchers potentially overcome one of the main objections to the SA of qualitative data - the problem of interpreting data originally collected by others.\(^4,5\) While familiarity with the context of data collection is an advantage, levels of familiarity can vary. For example, the data may have been collected by a team of researchers with split responsibilities for the study design, data collection, processing and analysis.\(^4\) Similarly, in the case of amplified analysis, where two or more data sets from independent secondary sources are combined and jointly analysed, each researcher concerned will have only partial knowledge of these data. Secondary analyses may also be co-conducted with researchers who were not involved in the primary study. In addition, as Mauthner et al. have pointed out, researchers’ ability to interpret their own data may also decline over time as memories wane; changes in researchers’ personal situation and new knowledge that they have gained since the primary study may also influence their re-interpretation of the data.\(^77\)
Only seven studies were found to use independent data sources. Of these, just three used archived material (and two of these used the same data set). Thus, Paget\(^{57}\) used archived data from Waitzen’s\(^{78}\) research and Weaver\(^{21}\) developed a ‘secondary ethnographic data analysis’ of Roth’s data derived from his study of tuberculosis sanitoriums.\(^{79}\) The latter data set was also used by Weaver and Atkinson to illustrate a book on the use of computer software to assist qualitative data analysis.\(^{22}\) Of the remaining studies, one was of naturalistic data in the form of lay autobiographies.\(^{34}\) In another case, the author was ‘invited’ by the primary researchers to conduct the SA.\(^{42}\) And in yet another case the data appeared to be obtained informally from another research team.\(^{71}\) Finally, one SA\(^{20}\) drew on a published analysis by Zussman,\(^{80}\) as opposed to the raw data itself.

There are several possible reasons why archived data has not thus far been utilised to a greater extent. Archives for qualitative data have not been established as long as those for quantitative data and the availability of data sets, although increasing, is relatively limited at present. In addition, there may be advantages to sharing data on a more informal basis, as in some of the amplified analyses identified, whereby the SA is performed by a team of researchers who can pool their respective knowledge of the primary studies from which the data were derived. The data concerned are also more easily shared on this basis, as compared with the relatively restricted access to data deposited in data archives and the associated costs involved.\(^{81}\) Unlike quantitative data which tends to be shared electronically and in processed form, qualitative data sets comprised of tapes and transcripts are often available for consultation in the archives only, hence it is less convenient to access this material and develop computer-assisted analyses. Finally, while the development of archives has begun to provide material for researchers to re-use, nevertheless, the methodology of SA of qualitative data has remained largely under-theorised, and hence researchers have lacked understanding of appropriate conceptual and methodological tools.

**Methodological issues**

In general, the appraisal of the conduct and therefore the quality of the studies was hampered by a lack of reporting on how these were actually accomplished. One of the best documented secondary analyses was by Szabo and Strang.\(^{65}\) Their report provides information on: the respective funding of the primary and secondary work; the relationship of each of the authors to the data; how informed consent and ethical approval for the SA was obtained; the composition of the data set; the use of grounded theory in the analysis; how the SA was performed; how the data were managed; and how the rigour of the analysis was established. However, many studies failed to report on key issues pertaining to the conduct of SA, including the selection of the data set(s) and sub-sample(s), how the ‘fit’ of the data and the
focus of the SA was ascertained, how the rigour of the analysis was established, and the possible limitations of the methodology.

In some studies, there was evidence that researchers had made efforts to establish the rigour of their analyses. The techniques employed included the establishment of audit trails, validation of findings using members of the study population, triangulation of data sources, consultation with the primary investigator(s), and independent coding of data. However, it was not clear from the reports how these techniques, most of which are used in primary research, were successfully transferred and applied in the context of SA.

Various methods, including grounded theory, narrative and content analysis, were applied in the studies. Grounded theory was the most popular approach, used in some form or other by several studies. In certain cases, this approach had been adapted specifically for use in the context of SA. Thus, one aspect of grounded theory - theoretical sampling - whereby the study sample and/or research topics are revised in order to purposively address emergent themes from the initial analysis, was deemed not possible because of the post hoc nature of the secondary studies. Hence, an alternative approach was developed, whereby theoretical sampling was applied within the available sample in order to compare and contrast the experiences of a given population. In addition, some complementary analyses effectively overcame the limitations of working with pre-existing data by conducting additional primary research to purposively address questions related to the focus of the SA.

**Ethical issues**

Only two studies reported seeking informed consent from the informants whose data was used for the SA. In one case this was obtained retrospectively and in the other prospectively. As the vast majority of studies failed to discuss the topic, it is not clear if consent was simply presumed or if it was sought but not reported. This finding raises the issue of whether informed consent should be obtained in all secondary analyses and, if so, how. A related concern is whether the sharing of data violates prior agreements concerning the confidentiality of the information. Existing professional codes of practice in the social sciences vary in the attention given to these issues in the context of SA. While they generally endorse data sharing in appropriate forms and stress the importance of obtaining informed consent, at the same time, it is recognised that there are some situations, such as covert research, where this requirement is waived. Researchers are, however, required to consider possible harm to subjects and, according to the American Sociological Association (ASA), to seek approval of institutional review boards (ethical committees) or relevant experts.
The evidence from this review suggests that it may be inappropriate to generalise about the need to obtain informed consent for secondary analyses, as this is likely to vary according to the characteristics of the study. For example, in the case of supplementary analyses by a lone researcher using auto-data, the purpose of this secondary study may be sufficiently related to the primary study to be covered by the terms under which consent was originally obtained; furthermore, in re-using her own data, the researcher is not breaking confidentiality. Conversely, it is doubtful whether consent could be presumed in the case of supra analysis using archived data.

Where informed consent for SA is considered necessary, obtaining this retrospectively may not be feasible. Research participants’ names may be unavailable because of confidentiality agreements in the primary study; in addition, subjects may have moved or died since they were involved in the primary research. Alternatively, consent may be obtained prospectively by primary researchers who prepare subjects for the possibility that the data may be stored in archives and/or re-used for other purposes. While this is the more practical option, there are two main problems with it. The first is anticipating and explaining to informants exactly how the data may be used in the future. Secondly, even where conditions are imposed on access to and use of data by others, control over usage of the information cannot be fully guaranteed.

Implications for the development of the methodology

Based on the findings of this review, the following nine key implications for the future development of the methodology have been identified. These concern the conceptualisation of SA (i), the reporting of SA (ii), the promotion of SA (iii-iv), and areas where further research is required (v-ix).

i) There is a need to better define the use of SA in reports and in electronic databases.

This review has confirmed Thorne’s observation that secondary analyses are not always clearly defined as such. The future classification and sub-classification of studies may be helped by this review which has shown how the methodology differs from primary analysis (in the form of selective reporting of research results), literature reviews and meta-analysis, as well as distinguishing six types of SA of qualitative data.

ii) There is a need more fully to document the ways in which SA was accomplished in reports of secondary studies.

This includes the basic accounting of how data set(s) and any sub-sample(s) were selected, whether or not consent was sought and obtained, how the fit of the secondary research question and the data were ascertained, how the rigour of the analysis was established, and
the limitations of the methodology. Reviewers and journal editors need to encourage authors to more fully report on the conduct of secondary qualitative studies.

iii) *The possibility of qualitative researchers’ extending their primary work through the SA of auto-data or through informal data sharing should be acknowledged and facilitated.* The main advantages of re-using these sources of data are convenience, researchers’ familiarity with the context of the data collection, and that researchers maintain ownership and control of the data.

iv) *Social science professional guidelines on the procurement of informed consent and preservation of confidentiality need to be revised so that they consider these issues more specifically in the context of secondary research.*

The guidelines of the ASA are presently the most comprehensive in this respect. In the UK, Qualidata, in conjunction with the ESRC, have also produced draft guidelines on the issues of confidentiality and copyright with respect to the collection of data and its possible re-use. These may form a useful basis for updating relevant professional guidelines.

v) *Further research is required on the use of SA of qualitative data in other subject areas.*

This review has focussed solely on the international health and social care literature. Other areas of research, such as education and crime, may provide evidence which will modify and/or clarify some of the conclusions arrived at in this study. Such reviews might also usefully consider the impact and utility of the methodology in social research, as it is developed over time and applied more extensively.

vi) *Further research is required on researchers’ attitudes toward the use of archived qualitative data.*

The lack of utilisation of this source of data raises issues about whether this is due to the availability and adequacy of current stock, practical difficulties in accessing and analysing the material, or principled resistance to the formal sharing of qualitative data.

vii) *Further research is required to establish what constitutes a worthy qualitative data set for archiving and re-use.*

This may be informed by related ongoing work to develop criteria for assessing the quality of qualitative research (led by Professor J Popay and funded by the Health Technology Assessment programme).

viii) *Further research is required on ‘how to do’ SA of qualitative data.*
Future consideration of the methodology could usefully examine the respective use of auto-and independent data sets (from formal and informal sources), as well as the different approaches to SA outlined in this review. It could also consider techniques for establishing rigour, and the use of methods such as grounded theory in the context of SA.

ix) Finally, there is a need for future research on the public understanding of social science research and, in particular, lay views on the re-use of data. This includes the circumstances in which informed consent should be sought for secondary research, as well as the meaning of confidentiality and appropriateness of related techniques for anonymising data.

Such research could usefully explore the claims of Qualidata\(^{(89)}\) that:

‘...most people do believe that research is for the public good and that their contribution will be used in some way to create a better informed society, and even go some way towards implementing policy changes’ (p.22).

In summary, this review has found that the SA of qualitative data has not yet been developed extensively but that different approaches have been pioneered which have potential for the conduct of original research using pre-existing data. In clarifying the conceptualisation of SA in relation to qualitative data and in highlighting areas for future research, it is hoped that this review will be instrumental in the future realisation of the potential of the methodology.

**ACTIVITIES**

The following activities have been undertaken or are arranged:

- ‘The possibility of secondary analysis’ Invited seminar presentation at the Epidemiology Department and Leicestershire Health Authority, University of Leicester. November 2000.
- Member of the Qualitative Methods Network, co-ordinated by Professor Jennie Popay, with proposed links to the Cochrane Methods Group and Campbell Collaboration.

**OUTPUTS**

- ‘Secondary analysis of qualitative data: An annotated bibliography’. See Appendix. This will be made available on the web.
• Contracted to publish a book based on the review with Sage (due December 2001).

IMPACTS
Appeals for information on the Internet for references to secondary studies also generated responses from researchers interested in the project, including researchers, students and archivists in the UK and abroad. Some studies which had not been self-defined as ‘secondary analyses’ in the publication were referred to me by the authors as possible examples. Other researchers whom I have contacted about their work have re-considered whether or not their work is best defined as ‘secondary analysis’. These initial responses suggest that the work, particularly when published in book form, will hopefully impact on how people conceptualise and define such studies in the future.

My involvement in the Qualitative Methods Network is primarily because of the implications of this work for: distinguishing secondary analysis from related methodologies and meta-analysis in particular; reporting the methods adopted in this form of qualitative research; and for developing methods for assessing the quality of qualitative work. While the Network is still in the process of defining its remit and links with other groups, it is likely that this will have a major international impact on the use of qualitative methods.

FUTURE RESEARCH PRIORITIES
See Results, points 5-9 (p.24-25).
REFERENCES


