TIMESCAPES METHODS GUIDES SERIES 2012

Guide No. 19

Qualitative Secondary Analysis: A Guide to Practice

Sarah Irwin & Mandy Winterton

INTRODUCTION

Qualitative secondary analysis allows researchers to generate evidence and insight from pre-existing qualitative data. There are many reasons why researchers might seek to re-use pre-existing data, for example revisiting contemporary or historical data with a set of new questions, exploring extant data as a preliminary to one's own study, or bringing extant data into conversation with primary or other evidence (see Irwin and Winterton 2011, Savage, 2005). Secondary qualitative analysis has a long pedigree although it is not very extensively practiced. Whilst not new, such data re-use has garnered increasing interest alongside the growth of digital archives, and the linked ability to make qualitative data accessible to re-users not involved in the original research project. There has been quite extensive discussion and debate about the epistemology and ethics of qualitative secondary analysis. We review some of this discussion elsewhere (Irwin and Winterton, 2011. See also Hammersley, 2009, Bishop, 2009). Whilst the broad consensus lies with the value and potential of secondary analysis it presents particular challenges. In this guide we discuss some of these challenges in re-using qualitative longitudinal data, and our strategies for tackling them.

KEY POINTS

For the kinds of analysis we report here, effective secondary analysis requires an in-depth understanding of the data set(s) to be used, in both practical and conceptual terms. Analysts need:

- a detailed understanding of the research project(s) including knowledge of the research aims, and design, and familiarity with project reports and publications.
- an understanding of the structure of project data, including knowledge of the ways in which it is embedded within, and reflects, the contexts in which it was produced (including, for example, knowledge of sampling decisions, and biases, recruitment strategies, and methodological tools and how they were put into use).
- a strategy for familiarising themselves with the project data as a whole, an understanding of the content of data, and its internal diversity, and a logic for subsampling for detailed analysis where a large volume of data prohibits a comprehensive reading.

There are different analytic strategies for working within, and where appropriate across, data sets. We have taken some conventional approaches and employed them as follows:

- Using a case based analysis we have sought to develop and refine concepts through building internal project comparisons;
- Through a strategy of translating concepts and evidence across research projects we have sought to enable meaningful 'conversation' across differently constituted data sets.
- Within one project we have used longitudinal case based analysis, organised strategically with reference to social diversity, to explore the interplay of specific influences on participants' (educational) identities and expectations, and how these evolve differently, through time and across social groups.

Developing a sufficient understanding of data, as a re-user, entails not only good descriptive knowledge of context, but a conceptual grasp of how data itself is embedded in the conditions in which it was produced. The methodological issues which arise here are challenging for all, not just secondary, analysts (although in our own practice, working across data sets put them into sharp relief).

We go on to briefly illustrate some different analytic strategies working within, and across, data sets and working with longitudinal data. We intend that this guide will offer insights for prospective secondary analysts and also hold relevance for primary researchers too, either those thinking of archiving data, or those who might find commonalities with our general interests in the links between data, evidence and conceptual development.

Timescapes has run a strand of secondary analysis activities, including various joint projects (e.g. collections of research and writing on common themes, and cross-project data sharing workshops). It has also included a dedicated Secondary Analysis Project, undertaken by the authors within the final two years of Timescapes funding. This project had a number of remits, including exploring the possibilities of secondary analysis within, and across, Timescapes projects, appraising the scope for extending the analytic reach of qualitative research beyond the boundaries of the original project designs, and exploring possibilities for making links with quantitative data. In this guide we draw on some of the lessons we have learned, seeking to distil some principles from our work in practice and, through the use of brief examples, we highlight some of the issues qualitative secondary analysts need to grapple with. We hope to offer some resources in the doing of secondary analysis. In this guide we follow a temporal logic, from getting to grips with a new data set to developing analyses and strategies for building understanding and explanation. We steer a course between the mundane 'doing' of things and conceptual abstraction, seeking to offer both practical insights and critical reflection.

Different archives, research projects and originating teams will have different protocols to be observed in undertaking data re-use and secondary analysis. Diverse forms of data, conditions of confidentiality and levels of access to data will render data differently amenable to re-use by others.

Prospective data re-users need to secure access to data, ascertain that originating researchers' and participants' permissions are in place, and observe ethical protocols and re-use procedures. These processes are explored in methods guides no. 16 and 18. Our focus here is on re-use and secondary analytic strategies and processes once data have been made available.

BACKGROUND
....

The Timescapes Secondary Analysis Project
The Secondary Analysis (SA) Project was designed as an integral part of Timescapes', to work with data generated by the primary research projects. These were all independently conceived projects, designed and run by different research teams, although they had a range of common concerns. In the SA project we explored the scope for researching links across data sets within, as well as beyond, Timescapes, and developing methodological insights. Part of the overall logic was to explore possibilities for enhancing the analytic reach of qualitative data sets. The timing of the SA project within Timescapes meant that it ran whilst some primary projects were still generating data. Most were still in the stages of preparing data for archiving, preparing data guides and undertaking their own analysis and writing. We were given access to part or (on occasion) all available individual project data. None of it was coded. We offer here not a full account of Timescapes-specific issues but draw on lessons which will have more general relevance. We do so through a grounded account which makes reference to examples of our research in practice. A more detailed extended guide is available as a working paper, along with details of our work, on the Timescapes website. (Irwin and Winterton, 2011)

RESEARCH DESIGN AND PRACTICE
....

Secondary Analysis in Practice
Whatever their varied purposes, secondary analysts will need to develop an in-depth practical and conceptual understanding of the data set(s) to be used as well as strategies for their analysis.

1. Getting Oriented to the Research Project(s)
Analysts will firstly need to orient themselves to the structure and content of their chosen project(s) and data sets. Metadata (data about data) provided by the originating project team, will aid in this task (see methods guide no. 16). The quality and extent of metadata supplied through social science archives is variable and Timescapes has sought to develop a particularly high standard for archived metadata. Secondary analysts then need to develop a detailed understanding of the project(s), grounded in the available data. A superficial understanding of data, or cherry picking of data or cases out of context, might allow (at best) a descriptive and partial account, and risks being misleading. Developing a detailed understanding of data sets requires getting to grips with:

timescapes
An ESRC Qualitative Longitudinal Study

- The structure of the project data. If this is a multi-method study, what types of data are available and how do they articulate with one another? What was the originating rationale?
- The structure of the sample. Analysts need to understand the sampling logic, the achieved sample structure and be aware of how the sample speaks to their own research questions.

2. Understanding Contexts of Data Production

Recognising the nature of data as contextually produced is important to effective qualitative research. For us, working across projects, as well as with longitudinal data, highlighted the contextual situatedness of all data. The significance of research design, methods, interviewers' interests as well as the impact of specific contextual factors are particularly visible. Much of the detailed manifestation of context will only be apparent through reading and interpretation of the available data. In our extended SA Guide, and in other project publications, we give detailed examples of the ways data is embedded in the contexts in which it is produced. Different dimensions of context include the proximate contexts in which research participants move and researcher-participant interactions. These impact on data and are part of the tacit knowledge of primary researchers. Also important are the contexts of the research project design, the ways participants are oriented to the research purposes, and the methods and modes of questioning used. Researchers need to maintain a critical awareness of how their research reveals particular facets of what are, often, multi-faceted experiences (cf. Mason 2002). For example, in one of our analyses we worked across data from two of the Timescapes projects: Work and Family Lives, and Men as Fathers. Within both projects, men were interviewed about aspects of their experiences and identities as fathers. However, in the different project contexts we saw examples of rather different kinds of accounts of fatherhood, even where particular lines of questioning were quite similar. It is important to understand how far these different accounts arise from different project contexts, where participants are oriented quite differently (for example, to the practicalities of managing family life and other commitments in Work and Family Lives, or to issues surrounding the social psychological dynamics and social identities of fatherhood in Men as Fathers).

Another illustration of the contextual embeddedness of data comes from answers to questions asked in common across the projects. These were a set of questions relating to age, biography and perceptions of generational membership. When we brought projects into comparison the task was especially revealing of the embeddedness of data in project contexts, and alerted us to some of the challenges of working across projects in a productive way. Even minor changes in question wording oriented participants differently, and these supposedly shared questions were managed in very different ways by project teams. (For concrete examples see Irwin, Bornat and

Winterton, 2011). Secondary analysts need to maintain a critical perspective on the workings of contextual specificity throughout their work. They need a nuanced understanding of the ways data is embedded in its project context, and need to deal with this when working across data sets or adducing evidence towards their own primary research.

3. Building an Understanding of the Data

Secondary analysts can expect to confront complex and often very large data sets in an archive. Commonly it may not be realistic for data re-users to read, let alone analyse, all the data produced during the course of the original research. They will need strategies for making inroads into this complexity.

In getting started on tackling the large volumes of data available to us, we followed different strategies. One was a broadly deductive strategy where we initially identified (participant) cases using summary indices. All Timescapes projects recorded some standardised socio-demographic 'base data' on their participants, allowing scope for subsampling according to given criteria. However, decisions based on this may have quite fundamental implications in shaping how a data set is read. As one might expect, interesting dimensions of diversity are not always accessible through summary indices, and one needs to exercise caution in using them. There may be times when a deductive strategy can be used for identifying theoretically interesting case studies to analyse in more depth. In familiarising ourselves with qualitative data sets this deductive strategy could serve only as an entry point. We then 'read outwards' across the data.

Another, more inductive approach meant that we initially targeted cases in a more ad hoc way, building up a more complete picture of project data by reaching outwards across cases. For example, we explored data on values, and expressions of identity and commitment, and extended our reading outwards from particular cases to gain a sense of how they fitted into the bigger picture. Such an approach is iterative. The first readings build a picture of diversity. In a very large qualitative data set, where it may be necessary to sample, analysts might seek to choose as wide a range of cases as possible, perhaps with reference to metadata if appropriate, or perhaps with reference to reading selected material within transcripts. This then becomes the basis for undertaking more detailed readings and analyses.

Although qualitative analysts often move between specific cases and theoretical generalisation, they build on knowledge of how cases are situated within their study data. Developing such an understanding is a challenge for both primary and secondary analysts. As secondary analysts, exploring internal diversity within project data (with reference to specific issues) and bringing cases into comparison was an important part of our approach. Such diversity may fall out in different ways according to the specific line of enquiry being followed. In writing up

research built on case based analyses it is important to offer clarity and transparency as to why particular cases are chosen to evidence an argument, and to show how they are situated relative to other cases, and how typical they are. Addressing such questions is good practice for all analysts, primary as well as secondary. However, for secondary analysts their distance from the data set, and from its originating rationales, makes this a particularly challenging task.

4. Developing Analytic Strategies

Data re-users will have many diverse analytic strategies. Here we give brief illustrations of our own analyses.

Example (a) Using a case based analysis we sought to develop and refine concepts through building internal project comparisons. As part of the original design of Timescapes, projects held a number of substantive themes in common, although the resulting data was often very different. As one example of proceeding in this way, we worked with data from one project (Work and Family Lives), read cases and inductively built an understanding of how a phenomenon of interest (experiences of time pressure) were manifested across the sample, and how individual cases were situated in this respect, and in relation to each other. Although 'inductive', our understanding connected to conceptual knowledge rooted in wider evidence and research (on gender, time pressure and work life stress). We refined our understanding by exploring women's and men's experiences of time pressure across different household contexts, for example comparing working women married to working men and working women whose partners took on quite extensive practical care commitments. This informed an analysis of the asymmetry of gendered experiences of time pressure across different contexts.

Example (b) Through a strategy of *translating* concepts and evidence across different research projects we sought to enable meaningful analytic 'conversation' across differently constituted data sets. For example, we brought the evidence from example (a) above into conversation with another data set, specifically Men as Fathers. Here we began with a more deductive way of entering into this project data set, starting with cases which looked potentially 'productive' with respect to our questions, based on project metadata (supplied to us by the team) about household divisions of labour. However, having followed this more deductive strategy we then read 'outwards', to give confidence we were interpreting the evidence in a way which was consistent with other cases/evidence within the project sample. We chose for in-depth analysis examples of men in diverse circumstances and took a particular interest in comparing men who were accepting of conventional divisions of labour and men who both worked extensively and desired extensive practical hands on care of their young children. That is, echoing the strategy for Work and Family Lives data, we sought to explore circumstances which might generate less conventional outcomes (lower

time pressure for working women in Work and Family Lives; greater time pressure for men in Men as Fathers). In effect we sought to bring evidence into comparison on the basis of translating our questions, and emergent hypotheses, to a new project context, since dissimilarity in project designs and samples meant that we could not simply ask identical questions across the two datasets. The analysis exemplifies a broader point, that secondary analysts need to be creative and critical in conceptualising how to translate between evidence held in different data sets.

Example (c) Through longitudinal case based analysis, organised with reference to social diversity, we explored the interplay of specific influences on participants' (educational) identities and expectations, and how these evolve differently, through time and across social groups. Working with data from the Young Lives and Times Study we undertook a case based analysis with reference to diversity within the qualitative sample as it mapped onto significant, class related, groupings across the population. We read much of the case material, including longitudinal interview data with young people growing from 14 to 17 and 18 years old. For the largest, middle class, group, with a university educated family background, we read enough cases to satisfy ourselves that we could identify participants who were typical in respect of our concerns here. We then selected for in-depth analysis a spread of cases chosen strategically to illuminate diversity in family background and resources, across non-graduate middle class and other class backgrounds. We analysed young people's temporal experiences of parental, school and peer influences in their evolving orientations to higher education (which was a majority expectation amongst the sample). Detailed longitudinal case based analysis may orient us to the particular, but it simultaneously reveals how the interplay of these factors over time varies by social background and circumstance, and provides a revealing lens on the temporal, biographical confluence of processes shaping class varying expectations. A case based longitudinal analysis organised with reference to how diverse (here class related) experiences are situated, and evolve over time, offers a resource in theorising the structuring of inequality. The resulting arguments about the shaping of diverse trajectories, amongst a numerically small and specific sample, could be tested and refined by exploring them across different contexts.

CONCLUSION

Secondary analysis is a challenging undertaking. It is time consuming. It requires great persistence in ensuring an adequate understanding of details which may be tacit for primary researchers. It can be frustrating for researchers to become 'users' when they may be more used to controlling the who, what and why of research design and data. In conjunction, it may be a risky course to follow in so far as outcomes are uncertain yet time commitments can be extensive. However, as many insightful secondary analyses stand testament, there is a depth of social scientific insight and progress which can be achieved. There are a range of reasons why researchers might seek to undertake secondary analysis, and with enhanced technology there is now an outstanding set of qualitative data resources readily available for exploration and analysis. We hope to have provided some helpful guidance and encouragement to would-be secondary analysts.

REFERENCES

Bishop, E. (2009.) 'Ethical sharing and re-use of qualitative data'. *Australian Journal of Social Issues* 44(3) pp 255-72.

Hammersley, M. (2009) 'Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues,' *Sociological Research Online* 15 (1) Available at: http://www.socresonline.org.uk/15/1/5.html

Irwin, S. (2009) Family contexts, norms and young people's orientations: researching diversity. *Journal of Youth Studies.* 12. (4). pp 337-54

Irwin, S and Winterton, M. (2011) 'Debates in qualitative secondary analysis: critical reflections' *Timescapes Working Paper 4*. Available at: http://www.timescapes.leeds.ac.uk/events-dissemination/publications.php

Irwin, S and Winterton, M. (2011b) 'Qualitative Secondary Analysis: An extended guide' *Timescapes Working Paper 7*. Available at: http://www.timescapes.leeds.ac.uk/events-dissemination/publications.php

Irwin, S, Bornat, J. and Winterton, M. (2012) 'Timescapes secondary analysis: comparison, context and working across data sets'. *Qualitative Research* 12(1) pp.66-80 (special edition: Advancing methods and infrastructure for qualitative longitudinal research in the UK: the Timescapes experience)

Savage, M. (2005). 'Revisiting classic qualitative studies'. *Forum: Qualitative Social Research*. 6(1) (art 31)

RESEARCH TEAM

Timescapes Secondary Analysis Project:
Dr. Sarah Irwin (Project Director)
Dr. Mandy Winterton (Research Fellow).

LEAD AUTHOR

Sarah Irwin, Sociology and Social Policy, University of Leeds
s.irwin@leeds.ac.uk