Recommendations for Secondary Analysis of Qualitative Data

Sheryl L. Chatfield

Kent State University - Kent Campus, schatfi1@kent.edu

Follow this and additional works at: https://nsuworks.nova.edu/tqr

Part of the Quantitative, Qualitative, Comparative, and Historical Methodologies Commons, and the Social Statistics Commons

Recommended APA Citation


This How To Article is brought to you for free and open access by the The Qualitative Report at NSUWorks. It has been accepted for inclusion in The Qualitative Report by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Recommendations for Secondary Analysis of Qualitative Data

Abstract
Publications and presentations resulting from secondary analysis of qualitative research are less common than similar efforts using quantitative secondary analysis, although online availability of high-quality qualitative data continues to increase. Advantages of secondary qualitative analysis include access to sometimes hard to reach participants; challenges include identifying data that are sufficient to respond to purposes beyond those the data were initially gathered to address. In this paper I offer an overview of secondary qualitative analysis processes and provide general recommendations for researchers to consider in planning and conducting qualitative secondary analysis. I also include a select list of data sources. Well-planned secondary qualitative analysis projects potentially reflect efficient use or reuse of resources and provide meaningful insights regarding a variety of subjects.

Keywords
Qualitative Research, Secondary Analysis, Online Research Data

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 International License.

This how to article is available in The Qualitative Report: https://nsuworks.nova.edu/tqr/vol25/iss3/17
Recommendations for Secondary Analysis of Qualitative Data

Sheryl L. Chatfield
Kent State University, Kent, Ohio, USA

Publications and presentations resulting from secondary analysis of qualitative research are less common than similar efforts using quantitative secondary analysis, although online availability of high-quality qualitative data continues to increase. Advantages of secondary qualitative analysis include access to sometimes hard to reach participants; challenges include identifying data that are sufficient to respond to purposes beyond those the data were initially gathered to address. In this paper I offer an overview of secondary qualitative analysis processes and provide general recommendations for researchers to consider in planning and conducting qualitative secondary analysis. I also include a select list of data sources. Well-planned secondary qualitative analysis projects potentially reflect efficient use or reuse of resources and provide meaningful insights regarding a variety of subjects. Keywords: Qualitative Research, Secondary Analysis, Online Research Data

Introduction

Quantitative researchers regularly conduct and publish results of secondary analyses of existing datasets including the National Health and Nutrition Examination Survey (NHANES) and the Behavior Risk Factor Surveillance System (BRFSS). Reports based on results of secondary analysis of qualitative data are far less common with the possible exception of meta-studies based on previously published research studies. Several factors probably limit the practice of secondary qualitative analysis when compared to secondary quantitative analysis. First, there is a lengthy history of availability and use or reuse of existing quantitative datasets, while in comparison, access to digitized records of qualitative data has increased more slowly. A related issue is that online access to text transcripts—often the preferred form for qualitative data analysis—might be limited due to lack of resources to transcribe audio or video-recorded interviews, or to create digital copies of sometimes fragile documents. Next, it is likely that more researchers are familiar with secondary quantitative analysis, due to early exposure from frequent use of these datasets in statistical analysis courses. Additionally, as with primary qualitative data, researchers may be dissuaded by the perceived and actual time-consuming nature of qualitative analysis. Lastly, quantitative datasets are typically de-identified, while qualitative data sources may provide a great deal of specific information about participants. Depending on the data source, researchers may or may not be provided with clear guidance regarding use of data in research and participant protection; this uncertainty may comprise an additional barrier to secondary qualitative research.

At present, a wide and likely growing range of pre-existing qualitative data sources are easy to access directly and indirectly via online sources. In a seminal work on qualitative secondary analysis, Heaton (2004) asserted that while use of conversation or discourse analysis techniques on existing data was not unusual, “there is no tradition of re-using data from qualitative studies” (p. 6). Although Heaton and others (e.g., Long-Sutehall, Sque, & Addington-Hall, 2010; Van den Berg, 2005) have advocated for re-use of qualitative data, there are still relatively few examples of published secondary analysis of qualitative data, and a limited number of sources that provide guidance on process alternatives. Therefore, the purpose
of this paper is to encourage researchers to use existing qualitative research data in thoughtful and creative ways.

Scholars reading this work are likely to have differing levels of experience or knowledge regarding qualitative secondary analysis, so I begin by describing general aims and typical types of data used for qualitative secondary analysis. I next identify advantages and acknowledge challenges associated with secondary analysis of qualitative data. In the next section of the paper, I provide my own typical process guidelines, and follow this information with four broad recommendations related to purpose, analysis, and selection of data. I also provide a list of some available data sources.

Overview of Qualitative Secondary Analysis

General Aims

According to Heaton (2004), secondary qualitative analysis might be undertaken for three purposes: to use the same data to explore other questions; to compare findings from additional analysis with findings from primary analysis of the same data for the purpose of “verification, refutation and refinement” (p. 9); to conduct a meta-study integrating prior research findings. There are clearly differences in type of data needed to address each purpose; the first two require primary data and the third requires research reports to be used as sources. In this paper, my focus is on use of open access data, and I emphasize the first Heaton’s purposes: use of preexisting data to explore new questions. This is also an aim that is similar to how most researchers approach the process of quantitative secondary analysis. While some of the same principles and practices I describe might apply to re-analysis of previously analyzed data, specific processes will depend on who is doing the re-analysis (i.e., the original researcher versus others) and aims of the re-analysis project.

Typical Data Types

Online open access qualitative data can be viewed as belonging to one of three categories: scholarly research data; credible current and historical textual or visual data that might be used for research; non-research data. Sources for the first two include libraries, archives and other repositories; sources for the third include social media platforms, any websites that allow comments or contributions, and discussion groups or forums. Based on my experience, individual interviews, gathered as part of oral history of an event or era, comprise the bulk of available online qualitative scholarly research data. Individual interviews may be available as audio or video recordings, typed transcripts, or provided in both formats. Other qualitative scholarly research data include unstructured, open response, and narrative information that might comprise a freestanding dataset or might be a part of a dataset that also includes fixed response survey items. Emerging scholarly data sources include repositories for researchers to store their own qualitative project data for the potential purpose of later verification; at present there are only limited data available through these sources. Credible current and historical text-based qualitative data that may be used for research include documents, such as diaries, letters, reports, news media items, and legal records such as wills or contracts. Visual data such as photographs and audio/video recordings of occurrences and events may also be available as current or archived qualitative data. Open access qualitative data also encompasses non-research data that are created and shared, on a limited or unlimited basis, for the aims of an individual or group, that might include communications, promotion of self, products, or issues, information solicitation, and other goals.
Scholarly research data sources should include information about participant informed consent and any conditions or limitations that apply to access or reuse of data. Current and historical qualitative data might or might not have associated limitations or conditions that apply to use. In the case of non-research data, such as social media posts, there are not likely to be formally stated limitations or conditions for use. Generally speaking, scholars may find it difficult to verify veracity of information when not provided by a scholarly or other official or authoritative source. The extent to which this is an issue depends on the goals of the research. For example, some researchers may wish to explore dissemination of incorrect or deceptive information.

Advantages and Challenges in Qualitative Secondary Analysis

Secondary analysis of qualitative research data has similar advantages to secondary quantitative analysis. These include time savings in the sampling, data processing and collection processes and ready availability of rich data for research projects that qualify for exempt status or expedited review by most institutional research review boards. Even when not entirely de-identified, data housed in open access archives have generally been gathered and made available through use of an informed consent process with participants, as suggested above, which may facilitate local review board processes.

Other benefits of secondary qualitative analysis include potential to access in-depth data provided by difficult to access participants, or data related to controversial or uncomfortable topics, reduction of burden on participants, and maximizing the value of participants' contributions to research through reuse of existing data when appropriate (Chew-Graham et al., 2012; Eastabrooks & Romyn, 1995; Long-Sutehall et al., 2010). An additional benefit is to be able to sometimes provide a voice to those whose thoughts and ideas were previously less valued or unheard due to cultural, social or political circumstances or priorities (Reilly, 2019). In some instances, less heard individuals were able to express privately themselves in contemporary documents, such as diaries or letters, or retrospectively in oral histories; qualitative secondary analysis can bring these individuals’ thoughts and experiences to life for a new generation of interested readers and researchers.

There are several practical and ethical challenges that might apply to secondary qualitative analysis. Hinds, Vogel, and Clarke-Steffen (1997) pointed out that there is a possibility that existing data will not be “amenable” (p. 411) to analysis for a new purpose or question. Bornat (2005) suggested researchers remain aware of the degree to which re-use of data might be perceived as deceptive, when the new research purpose is not identical to the original purpose participants were provided when granting informed consent. Thomson, Bzdel, Golden-Biddle, Reay and Estabrooks (2005) described variations in the process of removing references to people, places and things, to balance the risk of loss of confidentiality not only of participants, but of others as well, with the risk of unintentionally altering or misrepresenting the findings.

Clearly, researchers need to balance the benefits and potential costs of conducting any secondary analysis. I believe that the context specific nature of qualitative data presents different, although not necessarily greater risks than quantitative data. For example, generalized conclusions from larger data analysis might portray a group or region in an unattractive way whereas qualitative re-analyses may be more likely to impact smaller units including individuals, families, or communities. However, these are risks for use of primary data as well. My suggestion is that researchers who have reason to question the cost to benefit ratio of dissemination of findings from a particular secondary qualitative research study should seek guidance from the appropriate research review board, which typically includes community
members and institutional representatives, even if the nature of the data potentially make a given project exempt from official review.

**Process Guidelines**

I follow the same general processes, with minor modifications, to prepare primary data for secondary analysis that I use to prepare data for initial, content-focused analysis. These steps include formatting text for analysis, checking transcripts for accuracy, conducting data analysis, and creating and maintaining an audit trail. My priority when possible is to identify source data that includes both audio and typed transcript files. The former allows me to better engage with the data, and availability of the latter results in substantial time savings as much of the qualitative research I do involves working with a typed transcript. Although preparation and analysis of secondary qualitative data are time consuming, there might be potential time saved not only by beginning with a typed transcript but also through bypassing initial steps in the research process that might include participant recruitment, interview guide development, and conducting actual interviews.

**Clean up and format text.** Online qualitative data are presented in a variety of formats and file types including downloadable word processing or plain text documents, PDFs including scans of typed documents, or as embedded text on a web page. Data retrieval might be as simple as downloading and saving a file, or might require a copy/paste process, or use of some method to convert files. In my experience, most data retrieved via online sources, even when downloaded in a relatively intact form, require alterations in spacing, margins, paragraphs, and other document elements. Headers, footers, page, and section breaks may interfere with the readability and usability of the text. If using Microsoft Word, it may be very helpful to enable formatting marks to see and more easily modify or delete unwanted document elements. Other things I typically do include inserting continuous line numbers and use find/replace to change identifying information, such as inserting I for interviewer and P for participant. I also apply any special formatting such as insertion of columns or extra margin space as needed to write or type in codes or comments. You may need to do more or less to prepare your data, depending on your particular analysis strategy.

**Check the transcript for accuracy and begin pre-analysis.** During this step, I download or stream the audio or audiovisual file, when available, while using my cursor to follow the text file, word by word. If the participant speaks quickly or is difficult to understand for other reasons, I might load the audio file into transcription or audio engineering software, such as Audacity ® (Audacity Team, 2019), and play the file at a slightly slower speed. Although many interviews housed online have undergone professional transcription, I often find I make subtle corrections in the typed transcript. And, even though I conduct primarily content-focused research, as opposed to conversation or discourse analysis, I still like to hear inflections, emphasis, and other patterns of speech. I mentioned pre-analysis above because it is during this initial listen through that I might begin creating analytic memos, including those that might address my reactions to paralinguistic aspects such as tone of voice.

**Conduct data analysis.** Secondary data analysis is essentially the same as analysis of primary data; after all, any transcript is not the data itself but a type of translation. A cleanly formatted and accurate transcript document is most likely useful for any approach to analysis you prefer – from pencils and highlighters to importation into and use of specialized software.
Creating an audit trail and adhering to other quality elements. For any qualitative analysis, I typically save all ensuing versions of analysis files or documents, create, order, and save analytic memos and other process documentation, record decisions and discussions from collaborative projects, and retain other records that might include codebooks, process or theoretical diagrams and models, and other correspondence. One difference is that qualitative secondary analysis is not likely to yield anything like a participant contact list, or key to match names and assigned code numbers. In lieu of this information, I recommend you retain a master list of all of the relevant online information for each data source used, that includes web address and retrieval date. When useful, I suggest you consider taking and retaining screenshots of access or download pages, because websites and pages are not necessarily stable over time. Another aspect of quality control is compliance with permissions to use the data – both those associated with the data and those associated with governing institutional research board requirements. I think it is a good practice to inform the data owner of your intent to use the data in any dissemination via publication, presentation, or other means, even when this is not an explicit condition of use of the data. Despite recent changes to the Common Rule, some universities, including mine, require you obtain written verification that research is exempt from research board approval and oversight. This is not just a necessary part of research practice but also provides information that may be required before your work can be published.

Recommendations

As described in the previous section, although there are some differences in data access and data preparation, many of the processes involved in secondary analysis of qualitative research data are similar to those involved in primary analysis of qualitative data. Following I provide four broad recommendations that I hope provide additional guidance to encourage researchers to plan and conduct secondary qualitative analysis projects that might be of interest and use to them and to others with similar or overlapping interests.

1. **Have a focus/purpose in mind before you identify and review data and be prepared to refine it to develop meaningful research.** One challenge with any secondary analysis is the need to develop a question or purpose that is meaningful, that is worthwhile, and can be addressed using data that were not collected specifically with that purpose in mind. Many quantitative datasets are broad by design to allow researchers to address a variety of questions. Qualitative data, unlike quantitative data, does not lend itself to fishing expeditions for associations. However, for either qualitative or quantitative data, trying to derive a question from the data itself, without use of other supportive information is likely to result in a weak rationale for the project and associated challenges in writing a high-quality report. I suggest that matching purpose to existing qualitative data might require multiple stages of review of available data and review of existing research and might also require researchers spend some time refining their initial research question or purpose. Typically, exploratory, descriptive, and some process questions may be the most straightforward to address via secondary qualitative analysis when using oral history and other interview data. This is because oral histories are frequently planned for the purpose of documenting notable lives or events: therefore, associated interview guides are often developed to solicit information about an experience (e.g., living through a hurricane), a time period (e.g., being involved in the U.S. civil rights movement of the 1960s), or accomplishments (e.g., contributing to development of electronic medical records). Other types of online qualitative data may be more appropriate for other purposes or qualitative approaches.
2. **Use thoughtful sampling processes if the size of the data corpus or your method requires you to select a subsample.** Many sources for secondary qualitative analysis include a set of multiple related interviews or other types of data that address the same concern. In some instances, there may be far more interviews that most researchers would typically conduct, even in a large project. There are multiple ways to select a subsample. I do not recommend random sampling of qualitative sources as a preferred approach given that the point of randomness in selection or assignment is related to the process of probability sampling for generalizability to a population. In contrast, typical reasons for sampling in qualitative inquiry emphasize access to the most appropriate participants. Given this, I suggest a preferred strategy to select a subsample is to identify the sources that best address your purpose as respects their attributes, experiences, or features of the data itself. As an example, the Kent State University May 4 archives (Kent State University Library, 2019) were established to provide a central place to house information that related to the event on May 4, 1970: a student Vietnam ward protest that culminated in Ohio National Guard shooting into a crowd of students, killing four, and wounding several others. The archives include many one-on-one interviews with participants who have various roles and experiences and a large number of brief reflective passages submitted in writing by participants. The interviews tend to be richer than other data types, because they are often longer, and because interviewers were able to use prompts to elicit more detail. Using random sampling of either or both type of sources might reduce the size of the corpus of data but does little to ensure that a researcher is going to get the most relevant data to address his or her purpose. A researcher with an explicitly stated purpose might further refine his or her sample by focusing on role (e.g., Kent State University students versus other area residents), involvement (e.g., individuals who were on campus on May 4 versus those who were less directly involved in events) or other features of the data, the participant, and the experience.

   a. Less is less. To paraphrase Morse (2000), if you have thinner data, you will likely need a larger sample. If the online data you access is a series of responses to one or more open response items on a survey, you may need hundreds of sources to see much variation and find ample detail to conduct meaningful qualitative analysis.

   b. More is as much, or more. I suggest researchers resist the temptation to conduct analysis on all available interviews that address a given purpose, when all available interviews is a large number, unless ample time and resources are available and there is a compelling need to look at all cases. The same risks apply as with primary analysis, including failure to do deep analysis or become intimately familiar with sources, in an effort to work through the mass of data. Refer back to 2 above for considerations in selecting a sub-sample.

   c. As noted previously, one advantage of secondary analysis is the ability to uncover and present information from the less heard voices. I suggest seeking out data that reflects participants who are not the majority with respect to sex, income, social status, age, or other attributes, or social, political or philosophical views, to the extent this is or might be made consistent with the stated purpose.

3. **Quality, credibility, and risk of unintentional harm vary by data type.** I believe, in general, that willing participants who have provided informed consent for gathering and archiving their data are the participants with the most to contribute, and also those who are also most gratified by use of their contributions. Examples of these data include oral history interviews, testimonials, donated documents, and other textual or visual data sources. Historical documents, including letters, diaries, and other expressions,
that were not created for research purposes, have some advantage over modern participant-created data types in that there is less risk of causing inadvertent harm to living individuals, although living descendants may be impacted by results of research conducted on family members’ data. Depending on the nature of data, purpose and analysis, these results might be disturbing, or might be beneficial or gratifying, especially if previously unknown insights and abilities are revealed. Other online qualitative data, including non-research data, is used with least risk of inadvertent harm when participants have consented to its use in research, although published and intentionally disseminated data (i.e., newspaper reports; blog posts) might present less risk than casual social media posts that are sometimes created by people who have inconsistent understanding of privacy.

4. **Choose approach and analysis strategies wisely while keeping an open mind.** Some of the more flexible approaches to qualitative inquiry, such as case study, descriptive or generic designs, and mixed methods approaches, may be ideal for use with secondary data. Some of these encourage integration of other data types which might help counter deficiencies in some data sources. For instance, oral histories describing an event or era might be integrated with archived newspaper reports about the same event or time period. I also encourage researchers to keep an open mind in analysis methods. As an example, I have done some research with the US Centers for Disease Control and Prevention National Violent Death Reporting System (Centers for Disease Control and Prevention, 2019) data. The dataset consists of primarily quantitative categorical data, but includes some qualitative elements comprised of narrative case histories produced by law enforcement personnel and medical examiners. My initial inclination was to analyze the narratives simply by assigning categorical descriptors to reflect contextual details due to the varying quality and depth of the narrative sources. However, when I began a collaborative process of open coding on a subset of the larger sample, I was pleasantly surprised at the depth and nuance of our results that led to what I felt was an engaging and informative thematic presentation. Fortunately, to counter the overall thin nature of these narratives, as described previously in item 2.a. “Less is less,” we had a great many cases available.

**Sources**

Following I provide a select list of some sources for qualitative data that can be accessed remotely via the Internet, sometimes directly and sometimes by completing processes ranging from site registration to submitting a proposal or formal request for access. There are also many sources housed in libraries and not digitized but available for local use; I suggest you check with your institutional or public library to identify any available sources for secondary qualitative analysis. Because many organizations are continuing to collect data and digitize collections, I encourage interested researchers to continue periodically to search and identify new sources.

**Digitized Online Oral history archives**

- Kent State University: “Kent State shootings: Oral histories.” This digitized data archive includes recordings and transcripts for many interviews of individuals who reflect on the Kent State Shootings. There are other types of data available, including photographs and submitted reflections. [http://www.library.kent.edu/special-collections-and-archives/kent-state-shootings-oral-histories-0](http://www.library.kent.edu/special-collections-and-archives/kent-state-shootings-oral-histories-0)
• University of Michigan at Dearborn: “Voice/Vision Holocaust survivor oral history archive.” Included are audio recordings and written transcripts from Holocaust survivors.
  http://holocaust.umd.umich.edu/interviews.php
• United States Library of Congress: “Civil Rights History Project.” This digitized archive includes video recordings and transcripts of interviews with individuals who share recollections of the US civil rights movement during the 1960s.
  http://www.loc.gov/collection/civil-rights-history-project/about-this-collection/
• University of North Carolina at Chapel Hill: “Documenting the American South: Oral histories of the American South.” This extensive collection includes recordings and transcripts, subdivided into several categories.
  http://docsouth.unc.edu/sohp/
• University of South Florida Libraries: “Oral History Program (OHP).” Aims of this collection include use in research and instruction. The collection includes both audio recordings and transcripts, environmental studies, and sustainability, and Florida and local history are the specific focus areas of this resource.
  https://guides.lib.usf.edu/ohp

Limited Access Digitized Data with Available Qualitative and Quantitative Components

• Inter-university Consortium for Political and Social Research (ICPSR): “Stanford Civic Purpose Project: Longitudinal Study of Youth Civic Engagement in California, 2011-2013.” This project includes fixed response/survey results, available in a variety of formats, and transcripts of 50 individual interviews, available for on demand download by registered and authorized users. Access is typically limited to faculty, staff or students at one of the 776 member institutions.
  https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36561
• Inter-university Consortium for Political and Social Research (ICPSR): “National Firearms Survey, 1999.” Available files include fixed response/survey results and open item results available as a single file, with responses clustered by item. Files are available for on demand download by registered and authorized users
  https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4552

Restricted Access Data with Available Qualitative and Quantitative Components

• United States Centers for Disease Control and Prevention: “National Violent Death Reporting System (NVDRS) 2003-2016.” These data include a series of categorical descriptions of cases of homicide and suicide and text-based narrative reports derived from law enforcement and medical examiner reports. Cases reflect an increasing number of states, with all 50 US states reporting at least a portion of cases as of 2018. New data become available each fall with a two-year delay (i.e., in fall of 2020, the 2018 year data will become available). To access full/restricted access data, investigators must complete and submit a proposal, and the principle investigator must have an earned PhD. Once a proposal is accepted, CDC representatives will contact the investigator directly to arrange secure file exchange.
  https://www.cdc.gov/violenceprevention/datasources/nvdrs/RAD.html
  o NVDRS has an available public report system via WISQARS. Researchers can use WISQARS to examine trends by state, age, type of fatality, and other details, to help inform proposals to request restricted access data.
  https://www.cdc.gov/injury/wisqars/nvdrs.html
Discussion

Qualitative secondary analysis is not a new idea but perhaps use of this method will expand as researchers are increasingly able to access a range of data sources via ongoing data digitization efforts. Although there are many similarities with primary data analysis, including steps in data processing, analysis, and quality control considerations, unique challenges are presented in particular in matching data to purpose and purpose to value. Although there is substantial interest in secondary qualitative analysis of what I described as non-research data (i.e., social media posts), the spontaneous nature and ability to alter or hide aspects of context and identity, may serve to limit the credibility, and therefore the usefulness of these data. I believe the increasing availability of digitized data of the other two types I described, scholarly research data, and current and historical textual or visual data that may be used in research, warrants additional exploration, especially as archived data may have source and origin information that enhances credibility.

In conclusion, in the introduction to a 2005 special issue on secondary qualitative analysis, in the online journal Forum: Qualitative Social Research, Corti, Witzel, and Bishop observed the increasing availability of resources for secondary analysis and while asserting: “the need for more [resources] still exists, in particular for high quality and transparent exemplars of re-analysis” (www.qualitative-research.net/index.php/fqs/article/view/498/1073). Now, almost 15 years later, I believe there is still great need, as well as tremendous potential for scholarly contributions to understanding and insight on myriad topics, through thoughtful approaches to qualitative secondary analysis.

References


Author Note

Dr. Sheryl L. Chatfield, C.T.R.S., is Assistant Professor in the College of Public Health at Kent State University in Kent, Ohio. She received her PhD in Health and Kinesiology with emphasis in health behavior and promotion from the University of Mississippi in University, MS, and received her M.S. degree in recreational therapy from The University of Southern Mississippi in Hattiesburg, MS. Dr. Chatfield completed the Nova Southeastern Graduate Certificate in Qualitative Research in 2014 and is currently one of the Senior Editors of TQR. Correspondence regarding this article can be addressed directly to: schatfl1@kent.edu

Copyright 2020: Sheryl L. Chatfield and Nova Southeastern University.

Article Citation