5-2018

Evaluating Preexisting Qualitative Research Data for Secondary Analysis

Victoria Sherif
University of Kentucky, victoria.sherif@uky.edu

Click here to let us know how access to this document benefits you.

Follow this and additional works at: https://uknowledge.uky.edu/edl_facpub

Part of the Educational Assessment, Evaluation, and Research Commons, Quantitative, Qualitative, Comparative, and Historical Methodologies Commons, and the Statistics and Probability Commons

Repository Citation
https://uknowledge.uky.edu/edl_facpub/2

This Article is brought to you for free and open access by the Educational Leadership Studies at UKnowledge. It has been accepted for inclusion in Educational Leadership Studies Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Evaluating Preexisting Qualitative Research Data for Secondary Analysis

Notes/Citation Information
Published in Forum: Qualitative Social Research, v. 19, no. 2, article 7, p. 1-17.

Copyright © 2018 Victoria Sherif

This work is licensed under a Creative Commons Attribution 4.0 International License.

Digital Object Identifier (DOI)
https://doi.org/10.17169/fqs-19.2.2821

This article is available at UKnowledge: https://uknowledge.uky.edu/edl_facpub/2
Evaluating Preexisting Qualitative Research Data for Secondary Analysis

Victoria Sherif

Abstract: In this article, I explore the nature of secondary analysis and provide a brief history of the method. Qualitative secondary analysis is a relatively under-used method in education and the social sciences, often due to the lack of easily accessible, relevant, trustworthy, and complete data. I address some of the potentials and limitations that influence its use and explore criteria for assessing the quality and sufficiency of preexisting qualitative research data. Qualitative secondary analysis has important implications for qualitative researchers, students and practitioners interested in generating new knowledge via unobtrusive, reliable, valid, and time/cost effective research through the wider use of existing qualitative data.

Table of Contents

1. Introduction
2. Development and Purpose of Secondary Analysis
3. Advantages and Limitations of Qualitative Secondary Analysis
4. Evaluating Quality and Sufficiency of Qualitative Data for Secondary Analysis
5. Conclusion

References
Author
Citation

1. Introduction

In recent decades, secondary analysis has become popular among educational and social science researchers due to the increased quantity and accessibility of quantitative and qualitative data (BURSTEIN, 1978; DARGENTAS, 2006; FIELDING, 2000; HAKIM, 1982; HEATON, 2004; MURPHY & SCHLAERTH, 2010; VARTANIAN, 2011). These publicly available data provide researchers with resources to examine and analyze new hypotheses, inform research questions apart from the original objective of data collection, and derive new and/or additional interpretations and conclusions that were absent in the original research findings. Whichever research purpose is favored in a secondary analysis study, an important question arises: "Are preexisting data suitable, sufficient, and of a high enough quality to obtain new theoretical, empirical, and/or methodological understandings?" The objective of this article is to suggest evaluation criteria for the suitability, sufficiency, and quality of preexisting data employed for purposes of secondary analysis research. In the first sections, I outline the history and purpose of secondary analysis with a focus on the advantages and limitations of re-analysis of qualitative data. In the last section, I describe an assessment rubric to evaluate preexisting data's capacity to fulfill objectives of secondary research. [1]

This work is licensed under a Creative Commons Attribution 4.0 International License.
Forum Qualitative Sozialforschung / Forum: Qualitative Social Research (ISSN 1438-5627)
2. Development and Purpose of Secondary Analysis

Secondary analysis has a long tradition in the social, behavioral and education sciences, and has been an integral part of the development of research inquiry. In this sense, nearly 60 years ago, Seymour LIPSET and Reinhard BENDIX (1959) discussed an opportunity to re-analyze "existing data which were originally collected for other purposes" (p.xxvii). It is opined that the history of secondary analysis as a research strategy started in the last century before the Second World War and employed survey data. With the first national population census in 1790, followed by the collection of attitudinal surveys, extensive opportunities for secondary analysis were created (GLASER, 1963; SMITH, 2008). The original study conducted by Samuel STOUFFER and his team (1949) outlined the lives, relationships, attitudes, and adaptations of service personnel. Its later re-analysis has led to the development of theoretical propositions on race, class position, and social adjustment, as well as methodological insights on latent structure analysis and attitude scaling (GLASER, 1963; SMITH, 2008). [2]

During the 1970s, Janet ELASHOFF and Richard SNOW (1971) made an effort to re-analyze Robert ROSENTHAL and Lenore JACOSEN's (1968) study on the effect of favorable teacher expectations on student achievement during the first and second grade. The original study found that high teacher expectations toward student academic success and intellectual growth greatly positively correlated with student successful learning and performance. A re-analysis was largely prompted by the huge interest the findings generated, as well as questions concerning the accuracy of results based on the original data analysis, methodology, and statistical procedures (ROSENTHAL & RUBIN, 1971). Another example of the important re-examination of influential studies is Frederick MOSTELLER and Daniel MOYNIHAN's (1972) secondary analysis study, "On Equality of Educational Opportunity." They revisited findings generated by James COLEMAN and his team in 1966. Their secondary analysis identified inaccuracies, discrepancies, and issues associated with the sampling and measurement instruments, and thereby led to a questioning of the validity of the original findings (MILNER, 1972). [3]

With the growing popularity of re-examining preexisting survey data, the concept of qualitative secondary analysis appears to have first entered the literature in the 1960s, when Barney GLASER (1963) discussed the limiting nature of survey data in the light of secondary analysis: "The emphasis on survey data neglects other kinds of data, particularly field data, and hence limits the potential use of secondary analysis" (p.11). He posited that qualitative data could and should be used in secondary analysis, as a way to independently contribute to this growing body of knowledge (GLASER, 1963). [4]

Since the 1960s, numerous definitions of secondary analysis have appeared in the literature, many with subtle differences and emphases on its theoretical, practical, and/or methodological implications. One of the first definitions of secondary analysis is found in the scholarship of Gene GLASS (1976), who describes it as "the re-analysis of data for the purpose of answering the original
research question with better statistical techniques, or answering new questions with old data” (p.3). This understanding of secondary analysis for answering research questions divergent from the original ones and generating new conclusions and interpretations has been actively emphasized (GLEIT & GRAHAM, 1988; IRWIN & WINTERTON, 2011; LONG-SUTEHALL, SQUE & ADDINGTON-HALL, 2010). In 1978 BURSTEIN wrote about the capacity of secondary analysis "to contribute to knowledge because it has the potential to consider important questions without some of the limitations, or with a different set of limitations than those encountered in the original investigation” (p.7). A few years later, Catherine HAKIM (1982) comprehensively defined secondary analysis as "any further analysis of an existing dataset which presents interpretations, conclusions, or knowledge additional to, or different from, presented in the first report on the inquiry as a whole and its main results" (p.2). HAKIM suggested the use of secondary analysis was appropriate if secondary research aimed to identify additional indicators of an examined phenomenon; to reveal additional detail on the same research matter; to review a research matter from the perspective of a new theoretical framework not applied in the original study; or to answer research questions more thoroughly than in the original study through the application of a more sophisticated analytical methodology. [5]

HAKIM's definition of secondary analysis greatly contributes to the understanding of secondary analysis as a method to "generate new knowledge, new hypotheses, or supporting existing theories" (HINDS, VOGEL & CLARKE-STEFFEN, 1997, p.419). As an effective methodology that allows for exploring dimensions not explored in the primary study (DU PLESSIS & HUMAN, 2009), secondary analysis aims to uncover new meanings of information familiar to a researcher along with revisiting original research findings to assure their reliability and novelty (HEATON, 2004; KELDER, 2005; SAVAGE, 2005). Thus, the analysis of data collected by someone else (JOHNSTON, 2014) is common in research "pursuing a research interest that is distinct from that of the original work” (BUSTAMANTE-GAVINO, RATTANI & KHAN, 2012, p.36). This can be demonstrated by secondary analysis studies conducted by THOMPSON (2000) and RUSH, WATTS and STANBURY (2011). In the first case, THOMPSON examined three different datasets, one of which provided information regarding the extent to which family and work conditions before 1918 created economic and social difficulties in workers' earlier experiences. The study employed archived documented materials collected to investigate British trade unionism from the 1880s. The original data were collected by Beatrice and Sidney WEBB (1920), and included personal diaries and a large number of interviews accompanied by hand-written notes. Although the information was originally collected to record personal activity, politics and research work, it was suitable and sufficient to inform the history of the London labor movement in the 19th century. [6]

Recent research has also demonstrated the capacity of secondary analysis to generate valuable practical insights in the field of social research. For instance, Kathy RUSH, Wilda WATTS, and Janice STANBURY sought to understand the accounts of community-living older adults on adaptations they made in regard to mobility, and factors that influenced those adaptations. The secondary study was
based on the interview data and demographic questionnaires collected by Margarete SANDELOWSKI (2011) to examine older adults' perceptions associated with aging along with the experiences of weakness in daily life. The secondary analysis study found several adaptation techniques available to older adults such as selection, optimization, and compensation. Moreover, the findings also offered a basis to maximize mobility with aging as well as a framework to identify older adults’ motivations in overcoming mobility challenges, which had not existed in the previous literature. [7]

Another example that illustrates how re-analysis of archived data allowed for the generation of new theoretical knowledge and practical recommendations is found in Pamela KIDD, Ted SCHARF and Mark VEAZIE's study (1996). Conducted by the same researchers, the original study sought to identify safety decision making patterns of older farmers and determine what health risks they perceived and how they were prioritized in respect to risk avoidance behavior. Re-analysis of the original data revealed new perspectives on the relationship between occupational context and worker behavior. It also allowed the researchers to design injury prevention strategies, identify economic factors influencing farmers' safety decisions, and develop an educational module to prevent farming-related injuries. [8]

In addition to furthering theory and practice (TRZESNIEWSKI, DONELLAN & LUCAS, 2011), "the re-use of qualitative data provides an opportunity to study the raw materials of recent or earlier research to gain ... methodological insights" (CORTI & BISHOP, 2005, p.7). Recent studies have demonstrated how secondary analysis has informed method's assessments, ethical considerations in data re-use, and the use of various sources and types of secondary data (BISHOP, 2007; BRIASSOULIS, 2010; CORTI & THOMPSON, 2004; GOODWIN & O'CONNOR, 2009; LONG-SUTEHALL et al., 2011). While working on two different projects that focused on the re-use of interview data collected by Margot JEFFERY, Joanna BORNAT (2003) discovered methodological challenges associated with preparing data for archiving, data deposit and access, and level of anonymization. However, "the possibility that a researcher might find new evidence, or be able to draw new conclusions from archived data" (p.314) presents many opportunities to deeply engage in somebody else's data, actively learn from the original study and its researchers, when possible, and conduct research independently of any funding body obligation. [9]

The substantive examples presented earlier highlight a number of reasons for conducting secondary analysis including: examining new research questions with preexisting data sources (ANDREWS, HIGGINS, ANDREWS & LALOR, 2012; MITCHELL, 2015); developing knowledge that would have not otherwise been generated without secondary analysis (SALES et al., 2006); and expanding practical and methodological implications of the method (NOTZ, 2005; SALES, LICHTENWALTER & FEVOLA, 2006; SZABO & STRANG, 1997). Despite the fact that secondary data analysis has been employed for a long time, it has been used primarily in quantitative research, and its use with qualitative data is relatively recent (BISHOP, 2014; BISHOP & KUULA-LUMMI, 2017.). The most common purpose of qualitative secondary analysis, according to FIELDING
(2000), is to gain new insights by re-analyzing the data from new perspectives. Although there are many advantages of this method, practical and ethical drawbacks do exist (BISHOP, 2014; MASON, 2007). Taking them into consideration is crucial when evaluating the quality, suitability, and sufficiency of qualitative data for their re-use. Therefore, in the following sections, prior to suggesting criteria for the evaluation of qualitative data for the purpose of secondary analysis, some benefits and limitations of the method are described. [10]

3. Advantages and Limitations of Qualitative Secondary Analysis

The definitions of secondary analysis, and qualitative secondary analysis specifically, described in the previous section have been accompanied by an ongoing academic debate over the potentials and limitations of qualitative data re-analysis. This debate is focused on the possibilities and methodological, archiving, and ethical issues (HEATON, 2004; LICHTMAN, 2005; MAUTHNER, 2012; YARDLEY, WATTS, PEARSON & RICHARDSON, 2014). [11]

As Janet HEATON (2004) has pointed out, qualitative secondary analysis as a methodology "for conducting free-standing studies using preexisting data originally collected for other purposes" (p.9) expands on data collected via observations, interviews, and document reviews (KAUFMAN, GUERRA & PLATT, 2006). Qualitative data sources also include researcher’s notes, diaries, autobiographies, and open-ended questionnaires. Qualitative data used in secondary analysis research are often collected and archived as a product of independent qualitative studies conducted by a research team or independent investigators, as well as longitudinal research funded by private and government agencies (ANDRANOVICH & RIPOSA, 2012; DWORKIN, 2012; JACOBSON, HAMILTON & GALLOWAY, 1993; KAUFMAN et al., 2006; MAZZOCCHI, 2008). [12]

Secondary analysis of qualitative data can benefit the researcher intending to answer exploratory research questions or (re-)examine perceptions and experiences of a target audience (WINDLE, 2010). A more precise interpretation or emergence of new conceptual frameworks becomes possible, especially when the primary investigator conducts the secondary analysis inquiry themselves or serves as an advisor on the secondary research, such as with faculty and student projects (CORTI & BISHOP, 2005; CORTI & BLACKHOUSE, 2005; EBBINGHAUS, 2005; O’CONNOR & GOODWIN, 2010; PARRY & MAUTHNER, 2005; WINDLE, 2010). This is due to their extensive familiarity with the context of the original research. Qualitative secondary research, therefore, can generally broaden and deepen knowledge by stimulating a comprehensive understanding of the nature of an issue, especially when such issue is examined by the authors of preexisting data. It can reveal additional context for educational and social encounters and collaborations, as well as inform situated narratives around a particular topic (BROOM, CHESHIRE & EMMINSON, 2009). For example, in my doctoral dissertation (SHERIF, 2016), I re-analyzed longitudinal data from a sample of high school youth to determine practices of leadership development. The original study I conducted aimed to explore youth’s perceptions, motives, and attitudes toward leadership and its development. Results of secondary analysis
deepened the understanding of leadership development contexts as recounted by youth. [13]

Qualitative secondary analysis can also be beneficial in terms of economics. Reanalysis of already collected qualitative data saves time and money. The method is efficient because data collection is often a time consuming and expensive part of the research process (REW, KONIAK-GRIFFIN, LEWIS, MILES & O’SULLIVAN, 2000). Collection of data from various research sites and multiple sample subjects may be financially and time-restrictive as well. Many researchers, especially graduate students, have limited resources and are unable to cover indirect costs associated with obtaining access to specific populations for collection of specific data. Qualitative secondary analysis allows the researcher to overcome these types of data collection problems, thus, creating an equal opportunity for novice and other researchers to obtain and develop research independence, knowledge, and skills (SMITH, 2008). [14]

In addition to economic advantages, qualitative secondary analysis has social benefits. Research findings generated through this method can contribute to the specific body of knowledge without any intrusion into vulnerable populations (JOHNSTON, 2014; SMITH, 2008; ROBERTS, 1996). It has the advantage of not collecting additional data from individuals who require special treatment with respect to safeguards for their well-being and privacy (such as children, pregnant women, individuals with special needs, prisoners, etc.) or are challenging to recruit or access. Qualitative secondary analysis respects an individual's right to be left free from research inquiries about themselves or their activities/experiences (BULMER, 1979; SMITH, 2008). In cases in which samples may be difficult to recruit or hard to reach, qualitative secondary analysis can provide insights into sensitive issues, while protecting identities and privacy (REW et al., 2000). For example, Lenore BORIS (2015) aimed to explore experiences of HIV positive women in Kenya. Reanalysis of existing qualitative data enabled the investigator to share stories of these women on their "priorities and concerns post HIV diagnosis and how this diagnosis has reshaped their lives" (p.50), which otherwise would be challenging due to the remote location and limited access to this authentic population sample. [15]

Despite its apparent advantages, qualitative secondary analysis has its limitations. One of the major limitations, by definition, is that the data reflect insights collected for purposes different than those in the secondary research and may not adequately fulfill the research objectives of a new investigator (REW et al., 2000). The investigator is then challenged to explore new research questions derived directly from the data or shape the data so they would match the aims of the secondary research (HEATON, 2004). Therefore, qualitative secondary analysis appears to be useful as an extension of primary research, rather than for following up issues that emerged from the original study. [16]

Another major limitation of qualitative secondary analysis that expands on the issue of data fit is the completeness of preexisting data (HINDS et al., 1997; THORNE, 1998). The extent of data completeness is evaluated based on the
accessibility of the original context (MAUTHNER, ODETTE & BACKETT-MILBURN, 1998). The context of the original study is often referred to as background information, and includes knowledge about interactions between the researcher and study participants; their age, gender, social status, race, occupation, etc.; research settings, time, and place; sample selection decisions; data collection methods and procedures; and complete interview, observation, focus group, document analysis protocols, interview audio materials and their detailed transcripts (BORIS, 2015; IRWIN & WINTERTON, 2011). When the secondary analysis is conducted by one who did not collect the data (ANDREWS et al., 2012; CORTI & BISHOP, 2005; THORNE, 1998), it often becomes challenging to address data flaws and eliminate a misfit between the data and new research objectives (BISHOP, 2007; CORTI & THOMPSON, 2004; REW et al., 2000). In either case, "the lack of control over conceptualization of the study, data collection decisions, ... and/or biases in the original study that cannot be overcome in the secondary analysis" (BORIS, 2015, p.46) limits qualitative secondary analysis to only exploratory designs. [17]

Archived qualitative data are bound by time, which can make the process of accessing high quality data and interpreting them in the light of current issues challenging (REW et al., 2000). Archiving requires abundant human and time resources; researchers who usually work alone may lack the motivation, funding, time, and specific guidelines to assure high quality preservation of qualitative data (MAUTHNER et al., 1998). Moreover, systematic data organization and archiving is an ongoing process occurring simultaneously with data collection rather than an "add-on-task' at the end of the project" (HADFIELD, 2010, p.61). A more proactive approach enables the researcher to anticipate data archiving challenges and prepare data for archiving at various stages of material collection to assure protection of participant rights and identity. [18]

Archived qualitative data are also susceptible to time. Helene MORIARTY and her team (1999) argue that archived data may be outdated and lack information crucial to understanding research problems in the present day. Natasha MAUTHNER, Odette PARRY and Kathryn BACKETT-MILBURN (1998) note the potential discrepancy of theoretical and methodological lenses between existing data and secondary research questions. In this case, the data are inextricably connected to the unique settings of their collection. Thus, while it may be methodologically interesting to revisit archived data, "for the purpose of generating either new substantive findings or theories the data [are] wholly inadequate" (p.740). [19]

Ethical considerations for conducting qualitative secondary analysis include the process by which the subjects provided informed consent. Although secondary research implies no face-to-face involvement with study subjects and/or any intervention, there are challenges associated with confidentiality and subject agreement (CORTI & THOMPSON, 2004; HINDS et al., 1997; MOORE, 2007; THORNE, 1998). Archived qualitative studies embody an array of data about participants' backgrounds, knowledge, and experiences. Although datasets are typically cleared of identifying information, a secondary researcher may still
receive data violating anonymity guidelines. While conducting secondary research involving vulnerable populations, BORIS (2015) cleared secondary data from any identifying information to assure participants' confidentiality, as well as verified the accuracy of original informed consents accompanying the dataset. As she later wrote, the lack of informed consent and noncompliance of secondary data with research anonymity guidelines may become a risk to conducting secondary research. Similarly, MCLEOD and THOMSON (2009) in their book, "Revisiting Social Change," highlight confidentiality and participants' understanding of informed consent as cornerstones of a valuable, informative, and reusable qualitative data archive. Therefore, to address ethical challenges associated with the secondary analysis of qualitative data, researchers should be cognizant of the risks imposed by ethical considerations of the method and make an effort to verify the alignment of the primary research with research integrity guidelines. [20]

Still, although the secondary analysis of data is not without challenges, there are indeed fewer risks with the archiving and reuse of one's own data (HEATON, 2004). Conducting qualitative secondary analysis using one's own data generally avoids the limitations discussed above, and may present some additional benefits. For instance, Sarah IRWIN (2013) argues that revisiting data by the original data collectors can create a "form of critical distance." As a result, the primary researchers can critically evaluate the quality and efficiency of collected data from the perspective of new research questions and fill in the blanks in the original study background, data collection procedures, and missing information. The in-depth knowledge of the original research may also be beneficial in protecting participants' privacy and confidentiality, since primary researcher's immersion in the data enhances their ability to determine which information might be identifying (THORNE, 1998). Their familiarity and closeness with the data can, therefore, allow for eliminating epistemological and ethical concerns associated with the interpretation of data created by other researchers and participants, as well as more effectively assessing the fit between the original data and new research questions. [21]

4. Evaluating Quality and Sufficiency of Qualitative Data for Secondary Analysis

Current scholarship informs numerous examples of the re-use of qualitative data. However, there are only a few commentators on qualitative secondary analysis that seem to describe specific criteria for the evaluation of the quality and sufficiency of existing data. This section presents such criteria, which emerged as a result of careful examination of studies employing qualitative secondary analysis, as well as methodological, archiving, and ethical concerns. To identify the studies, a literature review of peer-reviewed publications was conducted. The search resulted in 40 sources that included doctoral dissertations, book chapters, and articles. The sources were identified through the following search terms: secondary data analysis, qualitative data analysis, secondary qualitative data, secondary data, and secondary qualitative data analysis. As the review of full text sources revealed, the process of evaluating the sufficiency and quality of existing
qualitative data includes evaluation of the fit and relevance of the original dataset to the present research, the background of primary research and collected data, dataset breadth and depth, data trustworthiness, and the timeliness of primary data. [22]

The process of evaluating preexisting data quality and sufficiency begins with conceptualization of a secondary study and selection of a research question (BORIS, 2015; DU PLESSIS & HUMAN, 2009; HINDS et al., 1997; JOHNSTON, 2014; KELDER, 2005; MEDJEDOVIC, 2011; MURPHY & SCHLAERTH, 2010; ROBERTS, 1996; SANDELOWSKI, 2011; SZABO & STRANG, 1997). In some cases, research can be conceptualized during examination of preexisting data with the research questions emerging directly from the data (GLADSTONE, VOLPE & BOYDELL, 2007; SHERIF, 2016). For example, "[t]here [should be] a logical link between the original data set and the question/s asked in the analysis, as the qualitative secondary analysis questions [arise] from the original data set" (DU PLESSIS & HUMAN, 2009, p.76). For instance, Brenda GLADSTONE’s team conducted a study to examine the subjective experience between motivation and schizophrenia. It aimed to explore psychosocial factors influencing motivation while living with schizophrenia, and resulted in sixty in-depth interviews with individuals diagnosed with schizophrenia and parents of ill family members. Descriptively rich data were later re-read and provided foundations for new research questions focusing on parents' help-seeking experiences. [23]

The quality of a primary dataset is determined by examining the breadth, depth, completeness, and accuracy of preexisting data (HINDS et al., 1997). The original data must allow the researcher conducting secondary analysis to understand examined processes, relationships, and subjective meanings (BORNAT, 2005). Only holistic and rich data can reveal the complexity of the original study and an examined issue (MILES & HUBERMAN, 1994). Such data, if re-examined, provide vivid and rich descriptions of a researched phenomenon while having a strong impact on the findings of the secondary study. The holistic and rich data also have the capacity to present every detail of the examined issue and illustrate how people perceive, react, and experience it. [24]

The breadth and depth of the primary data refer to the richness of data content, whereas completeness of the dataset includes the technical condition of the dataset and its accuracy. Specifically, having a complete dataset means there should be no or minimal missing or damaged documents (CORTI & BACKHOUSE, 2005; HARRIS, 2001; HINDS et al., 1997). The documents also should be accompanied by extensive description of methodologies employed to collect the data, field notes, study and sample design decisions, a detailed sample plan, and any additional information. [25]

Dataset quality is also addressed by examining data accuracy. Accurate transcription of interview and/or focus group data should have no or minimal typographical errors, incomplete sentences, and missing words (MacLEAN, MEYER & ESTABLE, 2004). Transcribed documents should be time stamped and accompanied by transcription protocols indicating instructions for transcribers.
and decisions that address inaudible text segments, overlapping speech, unfamiliar terminology, and language-specific nuances. An accurate dataset is also member-checked, which is completed during the data collection process and consists of the primary researcher restating, summarizing, and/or paraphrasing the information received from a study participant (GUBA & LINCOLN; 1981; KUZEL & LIKE, 1991). As part of this process, the respondent affirms that what was shared, heard or written down is recorded and understood correctly. Although dataset accuracy is critical to high-quality secondary analysis, assurance and compliance to accuracy standards require human and time resources (MAUTHNER et al., 1998).

The trustworthiness of preexisting data is often questioned by opponents of the method, perhaps because the reliability and validity cannot be addressed in the same manner as quantitative data (SHENTON, 2004). Nevertheless, several writers have demonstrated how researchers could address this issue. They note that comprehensive background information about the original study increases dataset quality and sufficiency (ANDREWS et al., 2012; ELLIOTT, 2015; FIELDING, 2000). Research background supports understanding and interpretation "because no data can be seen outside of a viewing context" (SANDELOWSKI, 2011, p.347). "The detailed contextual knowledge about the circumstances of the data collection possessed by the primary researcher" can foster both description and explanation (MITCHELL, 2015, p.3). In particular, knowing the research background allows for a complex understanding of nuances of the original study, as well as its theoretical and/or empirical contribution to the overall knowledge base.

How detailed should the description of the original study background be to make a judgment regarding dataset sufficiency and quality? According to Emmerentia DU PLESSIS and Susara HUMAN (2009), the original study background should include a description of research questions, the study population, sample selection choices, and the employed research methodology. An important question is how well the study settings and sample "of the original research project match those of the present project" (NOTZ, 2005, p.2). The settings of the original study and its sample must meet the expectations and needs of the secondary research (IRWIN, 2013; IRWIN & WINTERTON, 2011; JOHNSTON, 2014; NOTZ, 2005; WHITESIDE, MILLS & MCCALMAN, 2012). As IRWIN (2003) asserts, "in order to grasp the contexts underpinning diversity, and ... insights into conditions and causes" a thorough understanding and matching of study settings and people situated within the study sample become important (p.299).

Arguing that research background and collected data are related to each other, VAN DEN BERG (2005) also suggests evaluating background information about data collection methods and collected "raw" data (i.e., interviews, interview audiotapes, and transcripts); background characteristics of researcher/s and study participants; and information about the data collection site, time of data collection, and data collection settings. Additionally, Harry VAN DEN BERG and his colleague, Martyn HAMMERSLEY (2010), have emphasized the importance of accessing information about how study participants were selected and recruited.
and any other relevant information that may be helpful for a secondary data analyst to fully recreate the meaning and background of the original study. [29]

Evaluation of the quality and sufficiency of a qualitative dataset for secondary analysis also includes consideration of dataset timeliness and the possibility of further follow-up with the sample by the investigator of the initial study. Because the re-analysis is completed after the original study, data may be outdated due to changing theoretical and conceptual lenses, as well as historical events (LESKE, 1990; MORIARTY et al., 1999). If the original research had time-sensitive questions and aimed to portray historical events at the time of data collection, the secondary analysis may fail to provide an understanding of a research problem in the present day. Pamela HINDS and her colleagues (1997) note that a research problem interacts with and is affected by its context and that contexts change over time: "An analysis of a dated qualitative data set, in which the phenomena of interest may have since taken on different forms, is unlikely to provide a new or expanded understanding of the phenomena" (p.415). Therefore, it may be necessary for the secondary analyst to choose another more recent dataset or less time-sensitive data. [30]

The possibility of follow-up with primary study participants is especially important when the data quality is problematic, meaning data are outdated or missing (CORTI & BACKHOUSE, 2005; WHITESIDE et al., 2012). Although follow-up with subjects appears challenging due to numerous reasons, such as outdated or missing contact information, complete deidentification of study participants, original agreement that subjects will never be contacted again, etc., "in principle, it is wise to leave oneself the freedom to recontact on one's own behalf, or to ask informants whether they would be willing to be contacted for a different project" (THOMPSON, 2000, p.8). An opportunity to reconnect with subjects can be created when a primary researcher intentionally includes it in the consent form, sets up a flexible participation schedule of data collection, collects additional contact information, and/or tracks study participants during the study (WOOLARD et al., 2004). These procedures help not only facilitate the outcomes of a study, but also clarify characteristics of the data collection process and participant recruitment. [31]

A summary of these criteria for evaluating data for secondary analysis are found in the rubric below (Table 1). They allow for assessing: 1. the fit and relevance of preexisting qualitative data to secondary research; 2. the general quality of data; 3. the trustworthiness of the original dataset; and 4. dataset timeliness.

The rubric is designed as a diagnostic guide to identify criteria that could inform an assessment of the quality and suitability of preexisting data for secondary analysis. It helps remove the anxiety that comes with a researcher having to guess about what the dataset should or should not include, and makes it easier for one to justify the use of the preexisting data. If used for data archiving or
When using the rubric, one should note that each dataset is unique and has its own story. Evaluation of previously collected data is, therefore, the process of learning more about this story, and, specifically, about the original research, its background and settings, and participants' perspectives and how they were collected. When evaluation of preexisting data centers only on the rubric presented in this article, its results can lead to a risk of missing the overall meaning of the data. Furthermore, relying on the assessment rubric only, without taking into consideration the unique needs of the secondary analysis study, might turn the assessment process into a deficiency approach to secondary analysis, focusing on data deficiencies, limitations and errors. Therefore, the use of the rubric should be accompanied by critical evaluation of the general story presented by data, along with the consideration of secondary research questions and objectives. [34]

5. Conclusion

In their recent writing, Sarah IRWIN and Mandy WINTERTON (2011) suggest that an increasing capacity of research to generate abundant, rich, in-depth, and accessible data creates a profound opportunity to conduct efficient and cost-effective secondary inquiry. Asking additional research questions that have not been asked in the original study can reveal new knowledge, question current findings, or support existing scholarly findings and theories. The method also allows for the development, extension, and exploration of a phenomenon in a flexible and unobtrusive way. Particularly, the overall goal of this method is to contribute to the specific body of knowledge by providing an alternative perspective on topics without any intrusion into vulnerable populations (JOHNSTON, 2014; ROBERTS, 1996). As Barney GLASER (1963) suggests, "secondary data analysis can help save time, money, career, degrees, research interests, vitality and talent, self images and myriads of data from untimely, unnecessary and unfortunate loss" (p.14). [35]

In this article I have discussed advantages and limitations of qualitative secondary analysis. Yet successful qualitative secondary analysis is most effective when used with high-quality, relevant, rich, and complex datasets. As suggested in this article, careful examination and assessment of preexisting qualitative data can increase the validity and reliability of secondary analysis research findings. A systemic approach to the assessment of the quality and sufficiency of data used for secondary analysis can help qualitative researchers select an appropriate dataset and prepare already collected data for future archiving and re-use. [36]

To systematically evaluate the quality and sufficiency of preexisting qualitative data, an assessment rubric, presented in this article, describes in detail major parameters of data in respect to their fit and relevance for secondary research questions and aims; completeness, sufficiency, and accuracy; trustworthiness;
and timeliness. The rubric can be used as a tool to begin examining and learning about previously collected data. It allows for knowledge construction in regards to the original study and recognition of possible variations between the original and secondary analysis research questions. While the rubric provides evaluation criteria, which emerged from extensive literature review, it is theoretical in nature and needs to be practically tested. Overall, the rubric clarifies expectations for the quality and sufficiency of preexisting data, as well as saves time in data assessment and helps researchers think critically and broadly about research purpose, data context, and meaning. [37]

References


Broom. Alex; Cheshire, Lynda & Emminson, Michael (2009). Qualitative researchers’ understandings of their practice and the implications for data archiving and sharing. Sociology, 43(6), 1163-1179.


FQS http://www.qualitative-research.net/


Mauthner, Natasha S.; Parry, Odette & Backett-Milburn, Kathryn (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, 32(4), 733-745.


Moriarty, Helen J.; Deatrick, Janet A.; Mahon, Margaret M.; Feetham, Suzanne L.; Carroll, Ruth M.; Shepard, Margaret P. & Orsi, Allen J. (1999). Issues to consider when choosing and using large national databases for research of families. *Western Journal of Nursing Research, 21*(2), 143-153.


**Author**

*Victoria SHERIF, Ph.D.*, is a post-doctoral research associate in the Department of Educational Leadership Studies at the University of Kentucky, USA. Her research centers around issues related to qualitative secondary analysis. She is also interested in examination of youth leadership, its development and practice, as well as leadership of school superintendents to elevate their voices in various educational and socio-economic settings.

**Contact:**

Dr. Victoria Sherif

Educational Leadership Studies

University of Kentucky

111 Dickey Hall

Lexington, KY

USA 40502

E-mail: victoria.sherif@uky.edu

**Citation**