Secondary Qualitative Research Methodology Using Online Data within the

Context of Social Sciences

He-in Cheong, MEng

Imperial College London

Agnieszka Lyons, PhD

Queen Mary University of London

Robert Houghton, MSc, MSci

Imperial College London

Arnab Majumdar, PhD

Imperial College London

Author Note

Correspondence concerning this article should be addressed to He-in Cheong, Centre for
Transport Studies, Department of Civil and Environmental Engineering, Imperial College
London, SW7 2BU, London, UK. E-mail: he-in.cheong08@imperial.ac.uk. They are a PhD
candidate researching on people movement, especially those in situations of forced
migration.

Abstract

Qualitative research using interviews is a crucial and established inquiry method in social sciences to ensure that the study outputs represent the researched people and area rather than those who are researching. However, first hand primary data collection is not always possible, often due to external circumstances. Additionally, the use of secondary data, particularly open data, is progressively preferred to increase efficiency and gain geographical breadth. Therefore, this paper proposes a new step-by-step secondary qualitative inquiry methodology for online, publicly available interview data. Such procedural approach can help increase rigor, explicitly consider and mitigate potential pitfalls, and expand the research community's datasets. The 7-step methodology is based on a hybrid approach with elements of pragmatic qualitative approach, discursive grounded theory, and narrative approach and refers to research ethical principles of autonomy, equity, and diversity. One of the proposed steps is data quality assessment, which consists of quality assessment of data context and data content with a total of 16 quality dimensions to filter the collected data. The data analysis method comprises of content analysis for dataset categorization and thematic discourse analysis to answer the pre-set research questions. Additionally, the methodology covers the ethical and legal considerations relating to secondary online data and reporting the research findings of such data. By providing an example in the field of forced migration, we demonstrate how the methodology provides structure when conducting secondary qualitative research.

1   Introduction

Secondary qualitative data analysis can be a powerful method by which to gain insights that primary data analysis cannot offer. There is much literature using primary interview data, but often, the primary data represent either a small sample size or a limited regional

pool. Additionally, there often is a lack of continuity or connection between the different primary research in literature, which makes them difficult to combine. Furthermore, it is not always possible for various reasons, such as the safety of interviewees as well as the researchers, to conduct ethnography or first-hand interviews. Depending on the research question, a larger geographical and geopolitical coverage might be required, but there often is insufficient resource, budget, or time for first-hand data collection. To overcome these constraints, this paper proposes a method to use publicly available, online secondary data.

Databases of secondary qualitative data can be found in a variety of forms–whether one is already in a structured format or needs to be built via aggregating various sources. Analysis of such secondary qualitative data has established itself as a credible method for generating knowledge (Heaton, 2008), particularly in nursing (Szabo & Strang, 1997), as it removes the obstacle of first-hand data collection and its associated challenges of recruitment and the burden placed on both the interviewer and interviewee. There is a push in different fields of research to improve efficiency and to increase value for money; hence re-using existing data rather than generating new data is increasingly favored.

However, the use of secondary data has a number of potential limitations and their implications that need to be noted and be mitigated for (Chauvette, Schick-Makaroff, & Molzahn, 2019; Heaton, 2008; Hinds, Vogel, & Clarke-Steffen, 1997; Jacobson, Hamilton, & Galloway, 1993; Ruggiano & Perry, 2019; Sindin, 2017; Szabo & Strang, 1997). The key questions raised about the use of secondary qualitative data are related to data fitness, data quality, and limited clarity of the entire data collection procedure. These are in addition to ethical and legal implications of using secondary data (Chauvette et al., 2019).

This paper, therefore, presents a new step-by-step research methodology for using publicly available secondary data to mitigate the risks associated with using secondary qualitative data analysis. We set a clear distinction between overall research methodology and the data analysis method. The qualitative analysis method is only a small part of the entire qualitative research methodology. The methodology consists of applying pragmatic

qualitative approach with grounded theory and narrative approaches to collect, filter, and analyze secondary online data from multiple sources, which need to be filtered based on their quality. Some researchers argue against procedural approach of qualitative research because following a guide can "blindly" distract the researcher and prevent them from explicitly acknowledging their personal positionality and its influence on the research output (Savin-Baden & Howell, 2012). However, our argument is that without a procedural methodology supported by literature, secondary qualitative study can be daunting, difficult to approach, and lack rigor.

Our key contributions are the following:

1. A new step-by-step methodology for secondary qualitative research,

2. A novel data quality assessment based on qualitative context and content, and

3. A clear ethical and legal grounding for the research methodology.

The structure of the paper is of the following: The paper first discusses the ethical and legal considerations. Then it describes our overall qualitative research approach whilst highlighting its advantages compared to quantitative research. Then the concept of secondary data and its analysis is explained. This is followed by a step-by-step guideline of our methodology with each step in detail, providing a sample study in the field of forced migration.

Those in situations of forced migration or displacement are those who undergo involuntary migratory movement intra- and/or inter-nationally either induced by disaster or conflict (Migration Data Portal, 2022). Disasters typically refer to "natural" hazards, and conflict is typically referred to as a cause from humans, although this distinction is often blurry. In our case, forced migrants refer to refugees or those claiming refugee status and asylum because they have been forced to flee their home country because of 'a well-founded fear of persecution for reasons of race, religion, nationality, political opinion or membership in a particular social group' (United Nations General Assembly, 1951, p. 3). Regardless of the

driver of the migration, often forced migratory movement has to be individually arranged at the mercy of different people along their journey. Those in forced migration represent a highly vulnerable group of an extremely diverse background, who have experienced trauma in their original home country as well as on the journey in search of safety. This increases the difficulty of obtaining primary data due to reasons of logistics, safety, feasibility, language barriers, ethics, competence of the interviewer, and others. Using secondary data can connect the researcher with valuable data, but secondary research should have its own requirements and quality criteria to ensure rigor of the research. Particularly for a topic that is covered heavily in media, especially the quality criteria of the data and the trustworthiness of the findings need to be assessed explicitly to mitigate both the bias of the researcher and the primary interviewer of the data. The full case study of the application of the presented methodology can be found in the forthcoming paper by the same authors under the title of "Evaluation of assumptions of agent-based models of refugee movement based on ethical principles and secondary qualitative analysis."

2   Ethical Considerations

In the beginning of research, ethical as well as legal aspects should be considered prior to data collection because we, as scholars, have the moral obligation to do so and would also be required to submit ethical approval applications by research institutions. Yet, formally, using purely secondary text data is often exempt from requiring an application for ethical approval within research environments. However, there are still ethical issues associated with regards to retaining, sharing, and re-using secondary data as well as creating and archiving datasets from non-primary data (Heaton, 2008). On the other hand, if the data is freely available on the internet or in other published formats, such as books, permission for further use and analysis is implied when the ownership of the original data is acknowledged (Tripathy, 2013). Yet, during the use, confidentiality needs to be retained, and anonymization of the data needs to carried through from the primary research

(Thorne, 1998), especially if some primary researchers explicitly state that original names have been changed or omitted. When working with publicly available data that is fully accessible to anyone with uncensored internet connection, anonymization process follows what the original data providers do. In our research outputs, regardless of whether the interviewee was identified in primary study, we have decided not to report any name of individual participants and only provide relevant information about them with regards to the analysis. The ethical obligation to fully cite the reference of the data does not apply to revealing the full identity of individual participants.

Beyond the ethical implications of using secondary qualitative data, in the context of humanitarian crisis, continuous critical reflections are required to mitigate the unequal power between the researcher and the participant. The British Sociological Association (BSA)'s statement of ethical practice rests on active discussion and reflection (BSA, 2021). This will also improve the secondary qualitative analysis to reduce any biases or influence of personal values on the research outputs. Within the research field of forced migration, the following ethical guidelines are helpful: The Code of Ethics by the International Association for the Study of Forced Migration (IASFM) (IASFM, 2018), University Oxford's Central University Research Ethics Committee (CUREC)'s approved procedure on studies involving adult refugees in the United Kingdom (CUREC, 2020), Refugee Studies Centre (RSC)'s Researching forced migration: critical reflections on research ethics during fieldwork (Krause, 2017), RSC's ethical guidelines (RSC, 2007), Association of Social Anthropologists of the UK (ASA)'s Ethical Guidelines for Good Research Practice (ASA, 2021) and Internet Research: Ethical Guidelines 3.0 from Association of Internet Researchers (Franzke, Bechmann, Zimmer, Ess, & AOIR, 2020).

Whilst these existing guidelines focus on primary data collection, some of the principles can be transferred for methodology development, particularly those of IASFM (2018), as explained throughout this article. In addition, the updated 2021 version of ethical guidelines from ASA (2021) discusses the need to consider ethical implications relating to

use of published and online data. Overall, the researchers of secondary data equally need to evaluate the possible consequences of their work, just like investigators of primary data, and aim to do "no harm" through continuous self-reflection and critical approach to their own work. Different fields of research may have their relevant ethical guidelines on the analysis beyond the data management and usage.

3   Legal Aspects

Section 5.8 on "establishing rights to use published and online data" of ASA's Ethical Guidelines for Good Research Practice (ASA, 2021) state anthropologists should remain "sensitive" to potential consequences from ethical perspective as well as legal, in terms of intellectual property and copyright. There are exceptions to copyright beyond what is in the public domain. Public domain is defined by Erickson, Heald, Homberg, Kretschmer, and Mendis (2015) as works out of copyright. A work is considered out of copyright, if it has been over 70 years since the death of the author, if the work was never protected, or if it is used in a permitted way with a license. However, online works, even when publicly available in terms of accessibility, do not automatically fall under this category and thus are not exempt from copyright.

Analysis of online data itself does not infringe on copyright, as it is under private study, but the dissemination or presentation of the analysis, particularly when presenting direct quotations, needs further consideration of copyright laws. According to the Copyright, Designs and Patents Act in the UK Chapter III section 30 (Parliament of the United Kingdom, 2021) with the relevant amendment from 2014 (The Intellectual Property Office, 2014), use under "fair dealing" is permitted for non-commercial research in the context of criticism, review or quotation with sufficient acknowledgment. The use and amount of quotation would be determined by the purpose and what is necessary for the academic paper. There is a comparable Copyright Law of the United States Section 107 (United States Copyright Office, Library of Congress, 2021) that allows for "Fair Use" for purposes

such as criticism, comment or research for nonprofit educational purposes without infringement of copyright. Similarly, Directive of the European Parliament and of the Council 2019/790 Article 17(7) (Official Journal of the European Union, 2019) and its clarification (European Commission, 2021) and Directive 2001/29/EC Article 5(3) (Official Journal of the European Communities, 2001) allow for scientific research and quotations for the purposes such as criticism or review relating to a work or other subject-matter, which has already been lawfully made available to the public, with acknowledgment of the source and the author's name. Every researcher needs to consider the laws for their individual purposes, but the authors of this paper have sought the aforementioned articles. Although none of the legal documents define the term quotation, within this context we interpret it as "a phrase or short piece of writing taken from a longer work of literature, poetry, etc. or what someone else has said", as per the first definition of Cambridge Dictionary (Cambridge Dictionary, n.d.).

The mentioned laws legitimize reporting of data analysis of publicly available secondary data based on the following justifications: Firstly, presentation of discourse analysis on the selected limited quotations of online documents fall under the category of "criticism", "comment", and "review". Secondly, to meet the quotation criteria, we provide only the relevant part of the original source in an unmodified form. An exception to the modification has been clearly specified through use of square brackets [and]. Thirdly, we acknowledge the source, as per legal requirement, with a link to the original document and the known interviewer, author, and/or publisher in a tabular format with a unique identifier (ID). When presenting the quotation, if it is a sentence or shorter, we provide the quotation inline with the commentary in an identifiable format of foreign element through traditional textual means of double quotation marks " and ". If the length is beyond a sentence, we present the quotation through indentation and italics in the following format with the ID after the quotation, surrounded by commentary:

> *The quotation will be presented in this format. If there is an interviewer asking a relevant question, it will be highlighted in the following format:*
> [*Interviewer*]:
> *The question from interviewer*
> [*Interviewee*]:
> *Answer from the interviewee.*
> (ID-Example)

4   General Research Methodology

One of the purposes of qualitative research of interviews is to provide a pathway for co-production of knowledge, where the respondents' perspectives and interpretations are directly incorporated into research outcomes (Fedyuk & Zentai, 2018). In the process of developing the methodology, we have referred to Code of Ethics from International Association for the Study of Forced Migration (IASFM), particularly the three principles of "autonomy", "equity", and "diversity" (IASFM, 2018). These principles can apply to any social sciences research, as the ultimate aim is to "do no harm" or be non-maleficent, as per ASA (2021); BSA (2021); IASFM (2018). The principle of autonomy requires researchers to respect the researched people and to not position themselves as "experts" on the lives and experiences of the researched people. The principle of equity refers to the need for researchers to challenge unequal power relations between the researcher and the researched. The principle of diversity encourages researchers to include multitude of perspectives and experiences (IASFM, 2018). Although the Code of Ethics is predominantly about primary research (IASFM, 2018), we show that it is also applicable to secondary research throughout the entire article when describing the methodology.

4.1   Benefits of Qualitative over Quantitative Research

There are many benefits to qualitative over quantitative research, particularly in the context of forced migration. Quantitative data and the relevant methods cannot explain the motivations or the social meaning related to behaviors or how a phenomenon has

developed, whereas use of qualitative methods can help understand intentions (Castles, 2012). The study of forced migration, particularly by policy-makers, rely heavily on quantitative data and numerical modeling, as reviewed by Bijak (2022); Frydenlund and de Kock (2020); Klabunde and Willekens (2016); McAlpine, Kiss, Zimmerman, and Chalabi (2020). However, research about forced migration needs to have a refugee-centered approach (Clark-Kazak, 2021; Müller-Funk, 2020). Although quantitative data and its analysis have their benefits of offering clues to emerging concepts or answering high-level questions, to understand individual behaviors and have a research focused on individual person, using qualitative methods is more appropriate (Yin, 2018).

4.2 Our Qualitative Research Approach

To respect the three ethical principles of IASFM (2018) of diversity, equity, and autonomy, we developed a new hybrid secondary research methodology combining pragmatic qualitative research approach, discursive grounded theory, and narrative approach within the pragmatism research paradigm. In addition to adopting some of the existing overlaps between these three approaches, specific strengths from each approach are selected to develop the combined methodology. Although we apply it to a specific field, the methodology is designed to be data- and context-agnostic.

The research approach is based on the research philosophy of pragmatism. Pragmatism as a philosophy allows a flexible perspective to combine different research approaches to be pragmatic and practical where necessary (Savin-Baden & Howell, 2012). Pragmatism as a research philosophy and pragmatic qualitative research approach are not synonymous (Savin-Baden & Howell, 2012, p. 173). Although this research paradigm is the philosophical positioning of this research, when we refer to pragmatic research approach, we are referring to the pragmatic qualitative research approach, also known as "generic" (Merriam, S. B., 1998) or "basic" qualitative research (Merriam, S. B., 1998; Sandelowski, 2000, p. 335). Using this approach means using a combination of several methodologies or

having no commitment to a single methodology to identify recurrent patterns in the form of themes or categories for description, interpretation or understanding (Caelli, Ray, & Mill, 2003; Merriam, S. B., 1998).

We adopt a pragmatic qualitative research approach, as it allows for flexibility with a more relaxed data eligibility criteria to allow heterogeneous data types (Glasgow, 2013). This is particularly important for secondary data (Baldwin, Pingault, Schoeler, Sallis, & Munafò, 2022), where the sources may vary and the questioning of the interviews will naturally be inconsistent. Additionally, this research approach is suitable for meeting the ethical principle of diversity and feasible for a range of data sources. Because of the variance in the format of data, keeping a focused research aim is important, as per one the core elements of pragmatic approach.

In addition to pragmatic approach, which allows for a combination of methods where necessary, we draw on certain elements of grounded theory approach, although not in its completeness. The rigor of coding paradigm to develop theory grounded in the data provides the systematic yet flexible process that is suitable to secondary data analysis. Investigation of themes through a framework of open coding of discourse provides structure to analysis of interview data and answering the research question, as per discursive grounded theory (McCreaddie & Payne, 2010). The discursive nature of this research approach allows for the interviewee's story to be told, which is inline with the ethical principle of equity. Where we cannot use grounded theory approach as prescribed by Glaser and Strauss (1967) or Lincoln and Guba (1985) is the open-ended nature of the approach, as we are still using pragmatic approach. This means the element of sampling more and more data throughout the research to explore new theories (Szabo & Strang, 1997) will not be adopted for this methodology. When using the web as a portal to gaining data, it may first seem that there is an "infinite" amount of data, but quantity does not ensure appropriateness. Additionally, once available data is filtered based on its fitness, the suitable amount is likely to be limited.

Third relevant research approach is narrative approach, as it generally focuses on developing understanding of an experience through exploration and interpretation of discourse or personal stories (Leggo, 2011). According to Denzin (1997), all forms of qualitative writing is a narrative production, so the narrative research approach is relevant and applicable when analyzing interviews, even when obtained second-hand. Within the context of our research, the angle of narrative approach of learning about the interviewee and quote owner through discourse analysis of their words helps the research focus around the forced migrant and their experience rather than the researcher (Creswell & Poth, 2017). This respects the ethical principle of autonomy, as the narrative approach gives the quote owner the "ownership" of their story (Savin-Baden & Howell, 2012).

In summary, the proposed hybrid methodology is a pragmatic approach influenced by discursive grounded theory and narrative approach. Practically, this means carrying out discourse analysis of diverse set of narrative data based on pre-defined research questions whilst letting the data speak for itself by focusing on the individual quote owner or interviewee.

4.3   Secondary Data and Analysis

In the context of academic research and data analysis, primary data is referred to as data directly obtained via first-hand means, whereas secondary data refers to data collected by someone else for other purposes and use (Sindin, 2017).

Secondary data analysis refers to the analysis of such secondary data, which are pre-existing and is suitable for research of a question distinctly different from the original or primary study (Hinds et al., 1997). The analysis is often done by another researcher not related to the primary study using different analysis methods (Szabo & Strang, 1997). Secondary analysis in the context of quantitative data has been common for a while, but using secondary qualitative data has only really started in mid-1990s (Heaton, 2008). There is a growing push for re-using data due to the advantages of secondary data. The

general benefits are time savings, accessibility of large amount of data thanks to the internet, reduction of data collection cost, and widening the research focus in both longitudinal and geographical span, which can lead to new analysis (Sindin, 2017; Szabo & Strang, 1997). Particularly in research on vulnerable subjects, secondary analysis can reduce "respondent burden" as well as allow researchers to view dataset with a detachment that may be difficult to achieve during primary study (Szabo & Strang, 1997). Especially in fields where access to participants might be difficult particularly in controversial or uncomfortable fields, secondary analysis has even greater potential to maximize the value of participants' contribution to research (Chatfield, 2020). In the context of forced migration, our research suggests further advantages of minimizing stress added to the interviewee, as well as the interviewers, who may be psychologically affected by moral injury after having first hand listened to how interviewees were victims of acts that the interviewer considers immoral (Feinstein, Pavisian, & Storm, 2018). Therefore, using secondary data is inline with the ethical obligation of reducing "harm" (IASFM, 2018).

However, there are downsides of using secondary data and the pitfalls of associated secondary data analysis (Heaton, 2008; Hinds et al., 1997; Jacobson et al., 1993; O'Connor & Goodwin, 2010; Ruggiano & Perry, 2019; Sindin, 2017; Szabo & Strang, 1997; Thorne, 1998; Tripathy, 2013), as well as the ethical and legal issues outlined in sections 2 Ethical Considerations and 3 Legal Aspects. The disadvantages can be divided into two categories of data-related and methodology-associated.

The main three disadvantages related to the data are the following: 1) data fitness, 2) data quality, and 3) limited knowledge of data collection procedure. These disadvantages of secondary data, however, can be mitigated. Data collected for primary research with a particular question in mind may not fit or be appropriate for the secondary research question (Heaton, 2008; Sindin, 2017). Yet, if the data is relevant to the secondary research question although not in the ideal format, with flexibility of the methodology of the secondary analysis, this can be resolved. Even though secondary data quality cannot be

guaranteed due to lack of control over the data generation and content (Szabo & Strang, 1997), with an honest approach of outlining the unknown aspects of the data, extrapolation, exaggeration and misinterpretation can be minimized. Whilst using secondary data is often associated with limited knowledge of the data collection procedure and difficulties of "verification" of the data (Heaton, 2008) as well as limited "fidelity" of secondary data (Thorne, 1998), Heaton (2008) questions whether qualitative data can actually be ever verified, whether primary or secondary data. The concept of verification is derived from statistical and quantitative data and positivist-based approaches. Thus, for qualitative data, instead of verification, the "trustworthiness" of the data should be assessed (Lincoln & Guba, 1985).

In addition to data-related disadvantages, the main methodological challenge is related to data sampling (Hinds et al., 1997; Thorne, 1998). During secondary data sampling, because the process relies on existing research, the data pool could be disproportionately represented, and the assumptions made during original data collection could be distorted during the secondary analysis (Thorne, 1998). In addition, the original dataset(s) may not be sufficient to achieve saturation in case of application of grounded theory (Szabo & Strang, 1997) nor is a snowballing method possible. Secondary qualitative data can have a further limitation if the dataset comes from multiple sources and has inconsistent questioning or responses reflected in the observations (Hinds et al., 1997; Szabo & Strang, 1997). These challenges, however, can be mitigated through a rigorous methodology and appropriate research approach suited for the secondary data analysis with acknowledgment of limitations (Braun & Clarke, 2006; Lincoln & Guba, 1986), which are some of the main motivations for this article.

4.4 Our Choice of Data Types

The ideal choice of data type for qualitative research in our context is interviews, as the interviews contain personal stories. Interviews should not be seen to collect "hard facts"

but as a pathway for a researcher to gather information about the relevant field and insights into the lived experiences, knowledge, views, opinions and perspectives of individuals and the links between the individual and the collective (Fedyuk & Zentai, 2018, p. 174). Interviews are particularly valuable for obtaining implicit opinion, individual story or narrative through wording rather than "pure facts" (Fedyuk & Zentai, 2018; Kinsella, 2020).

Due to the advantages of secondary data analysis particularly given COVID-19 restrictions supported by our ambition to obtain narratives from different parts of the world, publicly available interviews or anecdotes from around the world were searched online. Because we are using secondary data, particular desired high-level format needs to be defined prior to beginning of the research. In the research of forced migration provided in this article, the ideal data types were interviews with and quotations from those who have been granted refugee status or were seeking it. An interview refers to a dialogue between an interviewer and the forced migrant. We refer to a quotation, within the context of this study, as a paragraph consisting of more than three sentences directly from the forced migrant and is presented in the form of written quotation, provided as an audio, or shown on video. Despite the inconsistencies in the questions, data lengths, and types, as long as the narrative of the forcibly displaced persons is visible to the authors, they were considered as valid data that would later undergo data quality assessment prior to data selection for analysis. Based on the autonomy and equity principles of IASFM (2018), only the text from the interviewee or the quote owner, and not the interaction, was justified as analyzable data.

This type of data, where we learn about the migrants' experiences directly from them, is the ideal data type, as we want to hear their story and how their journey has been so far. No one else can vouch or explain their journey as well as they can, and analyzing their own words give migrants autonomy (IASFM, 2018) in the research.

5   Step-by-step Guideline of our Research Methodology

In addition to the challenges of secondary research as mentioned in subsection 4.3 Secondary Data and Analysis, in current research realm of secondary analysis, there is a lack of rigor in the analysis and overall methodology (Ruggiano & Perry, 2019). This has the pitfall of possibly exaggerating the effects of researcher bias (Thorne, 1994, 1998). Bias of the primary researcher will naturally exist in the dataset. Therefore, when analyzing this dataset without deliberate and methodological attention to either primary or secondary researcher's influence on the findings, the primary researcher bias can be amplified particularly due to limited tacit understanding of the secondary researcher (Thorne, 1994). For this reason, we propose a new systematic step-by-step guideline with a set of methods for secondary data collection, filtering, and analysis to mitigate the downfalls of secondary data analysis, particularly in the setting of forced migration research when using online, publicly accessible data.

5.1   Step 1. Formulation of Research Questions

Setting a research aim is important regardless of whether the data is from a primary or secondary source (Taylor & Ussher, 2001). However, the type of aim is important to build the right setting for secondary qualitative research. This can be to supplement gaps in existing primary studies or in research areas, where direct data collection is not always possible. When working with secondary data, because the researcher has not developed a two-way connection with the participant directly, erroneous conclusions based on researcher bias are more likely than in primary research due to cognitive biases affecting interpretation of the data. This can be mitigated through a hypothesis-driven research plan, according to Baldwin et al. (2022). Some may argue that having research hypotheses can cause confirmation bias in the findings. Therefore, we use open-ended, non-binary research questions to define the aim of the study rather than hypotheses to increase flexibility in exploring themes that can help answer the questions. With research questions

in mind, the research is focused and the secondary data quality and fitness can be assessed according to the research questions. Having a research plan that is rigid yet flexible seem contradictory, but the rigidity refers to structure and flexibility refers to the approach. The authors propose a pragmatic research approach with elements of grounded theory and narrative approaches, as explained in subsection 4.2 Our Qualitative Research Approach based on specific research questions to draw on each methods' strengths and mediate the shortcomings.

A research question we have raised in our example case is, "Are refugees fleeing from their country of origin or fleeing to their destination?" The complete background to the formulation of the research question can be found in our other paper, but in summary, we wanted to understand the involuntary migratory movement in terms of directionality, irrelevant of the origin country nor the conflict. The question is not just about the logistics of the migratory journey but the effects of their situation on the mobility and aspirations of individual interviewee. Existing literature using primary data does not answer this specific research question (Bilecen & Lubbers, 2021; Bögner, 2005; D'Angelo, 2021; Forss, Mangrio, Leijon, Grahn, & Zdravkovic, 2021; Mangrio, Zdravkovic, & Carlson, 2018; Ryan & Dahinden, 2021; Sanchez et al., 2018), and obtaining primary data from those in situations of forced migration during the periods of a pandemic was not feasible in our case. Yet, with the hybrid research approach, having the perspective of an outsider to an existing conversation unrelated to the research question can be a strong advantage for revealing rich underlying information that may or may not have been available through direct questioning.

5.2   Step 2. Data Collection

Although Hinds et al. (1997) highlighted the need for selective data sampling due to qualitative methodologies being labor intensive and yielding large descriptive datasets, we have a different strategy. At first, data "dumping" from various websites and data sources

to local drive is needed to obtain a collection of data, regardless of their formats, as these will be filtered after anyway. The data collection method is loosely based on Jacobson et al. (1993) and follows these three steps: 1) Development of strategies to minimize selection bias, 2) definition of the target population and sample, and 3) application of criteria for data inclusion/exclusion prior to data filtering.

To minimize selection bias, using non-tracking search engine is necessary. We are not endorsing any search engine companies, but we used DuckDuckGo for data search. The reason for not using other major search engines such as Google is to avoid the "filter bubble" or the ideological frame of personalized search results to prevent intellectual isolation (Bozdag, 2013). For the example in this article, the authors used keyword "refugee", which represents the target population, and "interview", the desired data type, to obtain web search results of publicly available interviews with refugees or asylum seekers. For the example, no specific origin nor host countries were targeted to investigate a variety of geopolitical situations. We used a large variation sampling method, which is similar to maximum variation sampling without trying to reach every corners of the world, to follow one of the principles of code of ethics from IASFM (2018): Diversity. The variety helps recognize the diverse experiences of the forced migrants.

As a further search criteria for data inclusion during data collection, only English results were sought because the main language of research of the authors is English. Despite the limited explicit information on the original language of the interviews within the downloaded textual data, we used indirect indicators of the original language, such as the language of the original recording. Where no information is available, two different assumptions were made: Either the migrant spoke in English, or their words in another language were translated into English. The translation quality is assumed to have met the criteria of completeness, fluency, accuracy, verity, retention of meaning, and terminology (Lommel, Uszkoreit, & Burchardt, 2014; ISO 17100:2015, 2015), particularly with literal translations of pronouns of "I" and "we". However, regardless of whether the interview was

conducted in English or was translated, the authors accepted the data as they are. This distant approach to language translation may be seen as a limitation of the overall methodology, as one may question if the text is translated correctly. Yet whether translated or not, the superseding question would be if the data is credible and has not been contaminated with the translator's or the publisher's personal intention. Based on an ethnography study of a discussion between a Belgian lawyer and an Afghan asylum seeker without an interpreter in a "space of multilingualism" with linguistic challenges, Jacobs and Maryns (2021) show that if the interviewer has a strategy to approaching the asylum seeker's answers, the story no longer reflects the "narrative" of the refugee regardless of the original language. Thus, for our study, if the interview indicates that the publisher has a deliberative aim to reorient the narrative answers, the data is not valid based on lack of credibility regardless of whether the interview was translated. The assessment of the credibility of the source is explained in section 5.4 Step 4. Data Filtering.

The result from web search for publicly available, online, existing interviews or anecdotes from refugees consisted of interviews and anecdotes from refugees published by non-governmental organizations, charities, individuals, researchers, newspapers, to name some. The data often consists of an interviewer and the interviewee discussing the migratory journey of the interviewee at different lengths, styles, and levels of detail. All the data had been retrieved between September and November, 2021, and therefore, does not include any refugee following the Russian invasion of Ukraine starting from the 24th of February, 2022. Over 300 search results were downloaded at this step.

5.3   Step 3. Transcription, if Applicable

Some of the interviews may be in audio or video format. These need to be transcribed verbatim. When using a tool to transcribe automatically, the transcribed text needs to be checked against the actual interviews for accurate transcription. Whilst checking, we also added any interjections, exclamations, or non-lexical conversational sounds, that would not

have been picked up by transcription tools to add as much information on the transcribed interviews.

With regards to video data, in our case, consideration of body language and any interpretations on the interview setting have been prevented. Unlike primary researchers, secondary researchers will need to make assumptions on the interview environment based on their own personal thoughts because of the imbalance between limited information and abundance of personal opinion. To keep the data as close to reality as possible, contamination from researcher bias needs to be avoided, where possible. Hence, the transcription needs to only contain the verbal elements of the data.

In our case, any data not in written format were transcribed using the transcription tool, Otter.ai[1]. Post auto-transcription, the written data was cross-checked against the audio of the interview in its entirety. We added any non-lexical conversational sounds, e.g. "umm", or long pauses. We did not add any comments on the body language of the interviewee, as our methodology is based on discourse analysis on the text. Unlike primary research, secondary research needs to minimize interpretations subjective to interview setting as they would be based on perceptions and assumptions of the secondary researcher introducing bias.

5.4   Step 4. Data Filtering

During data filtering process, researcher bias needs to mitigated in two ways: 1) the data needs to be appropriate, 2) prior knowledge of data needs to be minimal, and 3) maintaining flexibility for the analysis (Baldwin et al., 2022). The content of the data does not have to directly answer the question, as the research is to qualitatively analyze and understand from the narratives of the interviewee, how they remember what they felt and experienced before and during their journey.

———

[1] We are not endorsing this tool, and there are many alternatives out there.

5.4.1 Data Quality Assessment.   Because the methodology involves secondary interview data, a systematic data quality assessment is required to "analytically filter" (Gibson & Brown, 2009) the available data to ensure only "eligible data" is included in the dataset for analysis. The answer to the question of what qualifies as "eligible" data relies on multiple dimensions and metrics.

From a scientific perspective, particularly when working with non-subjective, quantifiable and/or structured data, data quality is measured in dimensions and metrics (Laranjeiro, Soydemir, & Bernardino, 2015). 42 dimensions have been initially identified using Brackstone (1999); Chung, Nam, and Stefanone (2012); Dupuy (2012); Ehrlinger and Wöß (2022); Fox, Levitin, and Redman (1994); Laranjeiro et al. (2015); Mehrabi, Hassan, Sham, and Shahkat Ali (2009); Patton (2002); Pipino, Lee, and Wang (2002); West (1993), with the repeating dimensions of accessibility, accuracy, believability, completeness, consistency, interpretability, relevance, and timeliness. However, many of the dimensions are focused around quantitative or structured data. For instance, a key dimension in scientific research is "accuracy", which means a magnitude of error or closeness between data and the "real-world" (Ehrlinger & Wöß, 2022; Haegemans, Snoeck, & Lemahieu, 2016). The definition alone assumes quantifiable measurability within the data from a "truth" or "real-world". When working with interview or quotation data, the "real-world" is what the migrant perceives and does not lie within the researcher, as per ethics principle of autonomy (IASFM, 2018). Therefore, such dimension of "accuracy" is irrelevant.

Another dimension used for quantifiable data is "reliability" (Abdulla, Garrison, Salwen, & Driscoll, 2002; Chung et al., 2012; Patton, 2002). Reliability assesses consistency and repeatability of data, where reliability can be substantiated by other data sources. However, with subjective qualitative data, research aim is not to discover facts supported by evidence nor to carry out statistical assessment of similarities to report agreement amongst the migrants. The authors aimed to give each interview equal weighting for understanding individual experiences based on the ethics principle of autonomy (IASFM,

2018).

Even those dimensions that could qualify for qualitative data needed to be evaluated for applicability, as some are irrelevant to the context of forced migration. For example, the dimension "credential" (Dupuy, 2012) refer to the perceived level of acceptance by the researchers. However, this has been discarded, as having data only from sources known to the researchers can lead to bias within the database. The dimensions of "believability" (Laranjeiro et al., 2015; Mehrabi et al., 2009; Pipino et al., 2002) has been also excluded in data quality assessment, as it is not up to the authors to believe what the migrants have experienced or not, and every data is analyzed with respect to the migrants, as per ethics principle of equity (IASFM, 2018).

On the other hand, there are dimensions that are very much relevant. For example, "credibility" of the surrounding information needs to be assessed to ensure–to the best of the authors' ability–that the data may not have been fabricated nor have been contaminated nor fully co-constructed by the interviewer or publisher. In terms of secondary qualitative data, Hinds et al. (1997) provides a long list of assessment criteria for data quality and data fitness for secondary research questions, which the authors have summarized to the metric of "accessibility" and "relevancy".

Out of the identified dimensions, the chosen dimensions are categorized into data context quality and data content quality. Data context refers to the surrounding information, such as the interviewer questions, publisher credibility, and other peripheral data, and data content refers to the words from the migrants. As the study focuses on narratives from refugees and their words and not on the interaction between the interviewer and interviewee, the questions from the interviewer were used only as to provide context and develop metadata for categorization. Summary of the applied dimensions and their descriptions are shown in Table 1.

5.4.2   Sub-step 4a. Data Filtering - Data Context Quality Assessment and Sub-step 4b. Data Filtering - Data Content Quality Assessment.   Using the data

context and content quality assessment dimensions, an initial filtering system was generated

based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses

(PRISMA) (Moher, Liberati, Tetzlaff, & and, 2009; Page et al., 2021; Stewart et al., 2015)

methodology. PRISMA is originally developed to help reporting systemic reviews of

research or appraisals of published systematic reviews (Moher et al., 2009). Building upon

this method that is often otherwise used for literature review or clinical trials, we have

created a new data quality assessment procedure in form of a flowchart using the chosen

data quality dimensions as the elements to conduct data filtering. The intention of the data

elimination process is to only keep the data in desired format that is fit for the particular

study and to maintain the focus of the study. During qualitative study, it is not about the

quantity nor size of the data analyzed. Exclusion of data does not indicate that the author

is disbelieving or denying the narrative of the interviewees nor believing that the publishers

have malicious intent. The flowchart of our data filtering procedure is shown in Figure 1,

which shows the chronological order and the hierarchy of the dimensions listed in Table 1.

The process follows seven data filtering steps starting with identifying the data from search

engine results and initially screening the data based on information available about the

data, whether it is in the description or title. The next step prior to data retrieval is to

assess the website and organization for credibility. Some organizations may present

interviews carried out by someone not part of the organization, but this does not

necessarily mean that the source is not credible. To minimize biased sampling, we

consciously tried to disregard personal bias or historical knowledge relating to the

reputation of any organization.

After data is downloaded, where possible, the data format and the type of interview needs

to be assessed to see if they were an interview or quotation. Within the provided research

example, the interviewee needs to also have been granted refugee status or have been

seeking refugee status on the reasons stated in United Nations General Assembly (1951).

There were some cases where the interviewee is a voluntary and/or economic migrant or

someone who spoke on behalf of those in forced migration to perhaps protect the refugees' identity or ensure anonymity of a group of people. However, these interviews were excluded from analysis based on the ethics principle of autonomy (IASFM, 2018).

In some cases, data consisted of multimediality, which can support the credibility of the source yet can be perceived as "contaminated" data. The use of video footage or photos of the refugees or particularly hyperlinks to external references increases the credibility of the interviews. However, if other media forms were used to dictate certain reactions from the viewers, such as emotion-evoking background music in the interview video, these datasets are excluded for containing narrative elements not directly from the refugees. Furthermore, the desired video format is one not overly edited, where the statements from refugees are not sporadically presented and cut within a sentence to the point which the audience can lose the context of their narrative. One particular potential data contained a cartoon illustration in form of a satire with the interview to depict the situation of the interviewee. This source was excluded, as the story was being retold by the publisher with their own view presented in an art form.

Within the research, the dimension of interactivity of the inquirer excluded much data from the set. Only interviews, where dialogue or most of the quote originate from the refugee rather than the interviewer, were chosen as valid data for this study. Ideal format is a question from the interviewer with a paragraph of an answer from the interviewee. This is to respect the autonomy and equity (IASFM, 2018) of the forced migrants. In publicly accessible internet domain, much of the interview had highly edited text with a few words from the interviewee. Such interactive data may be relevant for other research questions, such as about co-construction of refugee identity, but for our research design, this was considered ineligible. We only wanted to include data where the interviewee or the owner of the quotation represented themselves as much as possible without interruption from the interviewer nor any overbearing reporting techniques. Most words of any text of conversation need to be from the refugees themselves, so that they have autonomy over

their own stories (IASFM, 2018).

Upon further assessment of data content, we realized some of the data was not relevant to our overall research topic, in our case forced migration and movement. For example, some educational institutions published some interviews about how interviewees were adapting at their own school. Because the intention of the interviews had particular purpose that may have had an influence on how the interview was run, these have been excluded. Once the data has been filtered to a set that was deemed appropriate for analysis, a further step is taken to assess the quality of the data content. In terms of (self-) representation, the interview was assessed whether the interviewee represented themselves or other people. One excluded interview was about how the interviewee believes to be the needs of other young people. Their answer is not explicitly based on their personal experience, which was a criterion for inclusion, and therefore, was excluded.

A short excerpt from accepted data that met all the data quality dimensions through the process in Figure 1 is shown below. In this example, a refugee from Somalia, who now lives in Australia, discussed his situation in Somalia when they fled. The full interview with the Australian Red Cross reveals that they left at the age of 15 unaccompanied:

[*Interviewer*]:

What was the situation like when you had to leave?

[*Interviewee*]:

I actually left Somalia not by choice, I just had to flee. My mother was away in Italy and my father was working the other side of Mogadishu and my sister was staying with another family. At soccer in the afternoon I heard this big explosion and then I rushed back to our house and that's the first time I'd seen the war began.

I couldn't find my family so I went with 60 people and joined in 300 people and we start walking towards Kenya. We didn't think we were going to Kenya but we just went that way. Without my parents. I was wearing the same shoes and the same pants and the shirt when I left Somalia so yeah, I didn't go back.

I was very scared, terrified, not knowing what was going to happen next. That's why I know what it's like, fleeing and having nothing with you. I was actually worried about my young sister, my sister two years younger. I feel guilty with that, going without her, but I couldn't cross the other side of Mogadishu. The only safe way to go was to go with this group.

(Australian Red Cross, 2015)

Beyond meeting all the data context quality dimensions, particularly the conformance of data type, key data content quality dimension that this excerpt shows is self-representation of the interviewee. The interviewee is talking about their experience with clarity, interpretability and understandability, which together form another key data content quality dimension.

Post data filtering, no data cleansing or pre-processing is to be carried out except for re-organizing quotations to combine text from the same speaker. As secondary qualitative data is often unstructured and consists of people's narratives, tempering with the raw quotes can impact the data quality and original intentions.

No mathematical formula is applied to set the desired sample size, nor is it feasible to use the concept of data saturation level with secondary qualitative data within the hybrid research method provided in subsection 4.2 Our Qualitative Research Approach. With "large" variation sampling method, there can be no end to learning something new with more data, so setting a set number for data collection is impossible. In addition, with

discourse analysis rather than corpus analysis, we are not searching for generalized conclusions supported by a number of agreed data but rather exploring different perspectives.

Using the data context and content quality dimensions and process, over 300 results have been narrowed down to 32 interviews and 52 quotations based on their data context. This has been further filtered to 19 interviews and 17 quotes based on their data content.

5.5   Step 5. Recognition of Researcher Positionality

According to Castles (2012), particularly a study on migration needs to include reflection methodology. Therefore, post creation of dataset and prior to data analysis, a positionality statement is recommended to understand the researcher's potential influence on data analysis and interpretations. Although an absolute removal of impact of researcher bias is perhaps never possible, explicit exploration of the researcher's "prospective reflection" (Boud, Cohen, & Walker, 1993) and personal stance on the research topic can help mitigate it and allow readers to come to conclusions about the research through transparency. Within the sample case study, the first author is most conscious of the fact that they do not have a refugee background and that their only understanding of the forced migration is from reading secondhand information, such as research papers, policy papers, news, and some personal discussions, which carry their own biases and will naturally influence the bias of the first author. Therefore, the first author has taken extra consideration to mitigate their own bias by ensuring not to make generalized conclusions, as per ethical principle of diversity (IASFM, 2018), and making a commitment to themselves to carry out the research by focusing on the data, as per the ethical principle of autonomy (IASFM, 2018).

5.6   Step 6. Data Analysis

After the data is filtered based on quality in step 4 as per subsection 5.4 Step 4. Data Filtering, the data analysis procedure follows the thematic discourse analysis method based on the pragmatic research approach developed by Braun and Clarke (2006). Although the

method from Braun and Clarke (2006) is developed for primary data in the context of psychology, elements of it still applies to secondary data analysis within the context of social sciences. The main difference between analyzing primary and secondary data is that in primary research, the interview questions may be directly related to the research question. In such case, consideration of the questions and the interaction between interviewer and interviewee may be important. Yet, with secondary research, because the questions may not align with the research questions, the focus of the secondary analysis is on the words of the interviewee.

5.6.1 Simultaneous to Step 6 - Continuous Reflection and Reflexivity.

Throughout the data analysis, continuous reflection and reflexivity is necessary. Reflection is the mental practice of thinking and meditating about the processes and products associated with the study (Savin-Baden & Howell, 2012), whereas reflexivity can be understood as examining or confessing one's own reaction through exploration of researcher-researched relationship (Finlay, 2002). A particular type of reflexivity, known as "reflexivity as discursive deconstruction" from Finlay (2002), is deemed to be a suitable pairing to discourse analysis to understand how language and text are used and how they impact the presentation. This can be done via a self-assessment journal during each sub-step of the analysis by explicitly stating the researchers' reflection and reflexivity. Through reflexivity, no single "comfortable" interpretation should be readily available, where the researchers settle on one interpretation immediately (Finlay, 2002). However, without moderation of the amount of reflexivity, over- or misinterpretation is more likely to happen at the cost of exhausting resources during research. The full process of writing the positionality and reflection statements in the form of self-assessment can be found in Finlay (2002).

5.6.2 Sub-step 6a. Data Analysis - Familiarization with Data.

This sub-step refers to familiarizing oneself with the data on a high-level without looking for specifics by reading through the data from beginning to end (Braun & Clarke, 2006), in addition to

knowing about the data from data filtering step. This is not the same as having prior knowledge of the data like a primary researcher would. The disadvantage of working with previously acquainted data is the element of subconscious bias on focusing on parts of data that have the strongest presence in researcher's memory. Therefore, where this step differs from working with primary data is that we ideally want to have a fresh look at the data. For primary researchers, because they have collected the data, they may have developed intimacy at naturally inconsistent level with different data. Whilst this personal connection to the data is considered an advantage for primary research to add richness to the analysis, for secondary research this inconsistency in familiarity is a disadvantage because it may lead to bias and favoritism of certain data.

5.6.3   Sub-step 6b. Data Analysis - Content Analysis and Categorization.

This sub-step is unique from other sub-steps of the data analysis process, as it aims to focus on "hard facts" relating to historical events to create a database for the dataset based on metadata. This is to give an overview of all the interviewees in a tabular format to check if variable sampling has been successful and if anymore data need to be collected to have a wider geographical coverage.

When using the content analysis method developed by Erlingsson and Brysiewicz (2017), the data–regardless of inconsistency in their formats–needs to be categorized to give structure to develop a database for the analysis. For the context of forced migration, we gave each data a unique ID and gave descriptions within the categories of the following:

1. ID of the interview data

2. Interviewee size, e.g. one person or a group of people

3. Sex of the interviewee

4. Age of the interviewee

5. Traveled alone?

6. (Estimated) Year of departure from origin country by the interviewee

7. Origin country of the interviewee

8. (Estimated) Year of arrival in the destination country by the interviewee

9. Destination country of the interviewee

10. Interviewer organization

11. Link of the interview data

The categories are all about the interviewee (except for the interviewer organization and link of the source of the interview) to focus on the refugees, as per ethical principle of autonomy (IASFM, 2018). The interviewer organization has been provided to meet the legal requirements of referencing the interview.

Sometimes there is no direct information in writing about each interviewee for all the categories. Therefore, content analysis is required to analyze the content of the interview to understand at least if the interviewee is an adult and roughly estimate what year they are referring to in their interviews based on the context of indication of war and/or other historical events. With regards to the travel size, particularly interviewees, who seemed to have traveled alone, did not explicitly state that they traveled alone, whilst others mentioned their travel companions or families.

For example, in an interview, from which the following excerpt stems, a Syrian refugee did not clearly nor directly said to the independent journalist whether he traveled alone or in a group. The interviewee described prior to the excerpt how he left Syria and then Turkey with his cousin. However, this was the only time throughout his interview, where he mentioned his cousin, and for the rest of the interview, he wrote in first person singular pronoun. After describing how he got smuggled into Greece from Turkey on a small boat with many other people, he explained what happened after he landed in Greece and walked to find the side of a cliff to sleep:

> *I had nine-thousand Euros sewed into my underwear and bagged in many layers of nylon. I had to check on them every now and then because I didn't know any of the guys I was walking with, but later I got to realize they were decent folks who were also afraid for their own money.*
>
> (Azoulay, 2018)

The interviewee explicitly says that he didn't know anybody he walked with whilst describing everything in first-person singular pronoun. This indicates that he may indeed have been traveling alone without pre-arranged travel companion based on the content.

5.6.4   Sub-step 6c. Data Analysis - Open Coding.   This sub-step forms the baseline of thematic discourse analysis with elements of discursive grounded theory (McCreaddie & Payne, 2010) and narrative research approach (Creswell & Poth, 2017) to code pieces of data specific to individual experiences without needing to directly relate to the research questions (Braun & Clarke, 2006). Initial codes in secondary analysis is purely based on the text from the interviewee, and there is lack of consideration of other aspects. Otherwise, the method is similar for primary and secondary analysis.

In this sub-step, the data is analyzed by categorizing data text into initial codes by letting the data speak for itself, based on discursive grounded theory. The codes are personal to the researcher, as they are based on how the researcher reads the text. Therefore, this stage is often not presented in literature.

5.6.5   Sub-step 6d. Data Analysis - Axial Coding for Initial Themes.   Following the open coding process, the initial codes are grouped into axial codes, similar to when undertaking primary data analysis. The procedures follows the recommendations of McCreaddie and Payne (2010) and Braun and Clarke (2006), where the initial codes are grouped and laid out to detect any patterns. Whilst the research questions should still be at the back of the mind of the researcher, the element of grounded theory of letting the data speak for itself is important to mitigate bias and reduce the influence of the researcher during the coding process. When combining initial codes to a single axial code, always

referring back to the data rather than just the code headings leads to findings of higher confirmability. The axial codes do not directly have to relate to the research questions.

5.6.6 Sub-step 6e. Data Analysis - Developing Final Themes.

At this sub-step the research questions are re-introduced to develop themes out of axial codes directly relating to the aim of the research. Overlapping of themes needs to be avoided, but same codes being grouped to multiple themes is allowed. Development of an initial thematic map can help this process (Braun & Clarke, 2006). Themes are developed via an iterative trial and error process of testing different possible themes to group codes. Always referring back to the original data and not just the codes can give higher trustworthiness of the result.

5.6.7 Sub-step 6f. Data Analysis - Answering the Research Questions.

At this stage, the themes need to not only relate to the research questions but also directly answer them. When selecting the relevant data or quote to report the result, the ethical principle of diversity of IASFM (2018) needs to be addressed to avoid repetition or concentrated representation of one demographic over another. The researcher's process of considering diversity can also reduce researcher bias to always cover a range.

In the following example, the refugee from Afghanistan who now lives in the state of New South Wales (NSW) in Australia, told the interviewee that she fled Afghanistan:

> [*Interviewee*]:
> *There was a lot of fighting, bombings everywhere and so we had to immediately leave.*
> [*Interviewee*]:
> *We lost our parents over there and I had to run with my three brothers and one sister.*
> [*Interviewee*]:
> *In Afghanistan or Pakistan if the Taliban gets you, if they catch that you're not wearing a burqa, something on your head they're just going to take you. They're going to tell you to sit down and they're just going to shoot you in the head and bang.*
> (State of New South Wales (Department of Education), 2020)

Returning to the example research question of "Are refugees fleeing from their country of

origin or fleeing to their destination?", in this particular case, the interviewee is expressing that she and her siblings "had to immediately leave" following bombings. There is no indication of where they were fleeing to, and the focus on this excerpt is that they were fleeing from Afghanistan. Although one may question if she is representing herself–based on one of the data content quality dimensions–when she is talking about how the Taliban treated people, the excerpt is still representative of her view and perception of Taliban and shows how she felt threatened as a woman in Afghanistan. The research question can be answered without the text from the interviewer, which is irrelevant for understanding whether she fled from Afghanistan or to Australia. This shows that secondary research can be done, even if the interview questions are not related to the research question.

5.7   Step 7. Assessment of Trustworthiness of the Findings

Post analysis, the trustworthiness of findings (Lincoln & Guba, 1985) needs to be assessed with focus on the criteria of dependability, credibility, and transferability based on the researcher's personal stance and positionality. Expecting a researcher to be 100% consistent throughout the entire process of qualitative analysis is unrealistic because different experiences and environments can influence research even when working with exactly the same dataset. Also, it is improbably for two researchers, who will naturally have different personal stances, to have the same findings. Therefore, rather than using the traditional assessment method of interrater or intrarater reliability, within the context of naturalistic inquiry of inconsistent secondary qualitative data, we assess if the findings are trustworthy based on the data and the researcher's report on self-assessment. This means that the findings need to be self-assessed on whether the researcher's bias, personal stance, and previous knowledge of the field influenced the analysis.

When using the proposed hybrid research method, the analysis and findings should reveal almost nothing about the researcher's personal stance. For example, the first author has their own views about human rights, and some of the data revealed that the interviewees

witnessed other refugees dying on their journey. However, the findings, (the result of the analysis) cannot show how that made the researcher feel. Not letting the personal stance influence the analysis is crucial in secondary research because bias needs to be managed and mitigated where possible when working with data with unknown variables, such as interview setting, interviewer-interviewee relationship, and sometimes even the questions. Because the secondary researcher was not present during the interview, any deliberate personal influence on the analysis will make the result unbalanced and untrustworthy.

5.8   Summary of the Steps of the Secondary Qualitative Analysis

The overall proposed method for data analysis is shown in Table 2.

6   Discussion

The proposed methodology of a step-by-step guide on secondary qualitative data analysis is not to be considered as a checklist exercise but rather as guidance. This may be particularly useful for those making a transition from quantitative to qualitative analysis and requiring a starting point.

6.1   Self-reflection

The authors have dedicated a paper specifically to research methodology because they believe the two fields of qualitative and quantitative research could be bridged through an explicitly written methodology that any systematic thinker can understand, regardless of background.

The first author is of a positivist research background who has previously found much comfort in quantitative analysis, and therefore, the journey involving entry into the field of contemporary social research has had its personal challenges. Particularly in the beginning, the self-resistance to accepting other research methods and interrogating the definitions of "truth" or "fact" proved to be an obstacle to reaching the stage of understanding social and qualitative research beyond corpus analysis. Overcoming this resistance came at the

effort of consciously questioning or even sometimes doubting previous education methods and beliefs of science. It would have not happened with the help of the co-authors, who have linguistics, psychology, and social science backgrounds. In retrospect, this journey perhaps could have been easier, if there were more research reflecting on differences between the philosophies of qualitative and quantitative analysis.

6.2   Limitations

As outlined in the section 5.2 Step 2. Data Collection, the main limitation related to using online secondary data ultimately narrows down to not knowing every step of how the data was collected first hand. Where there is no video or audio of the interview, the completeness and accuracy of transcription cannot be checked nor can the interview setting and condition be known. The researchers can only assume that the presented data is transcribed verbatim based on the credibility of the organization or the provider of the data, which is one of the dimensions of data content quality. As for the quality of interviewing method, this can only be inferred from the interaction between the interviewer and interviewee. This is covered by the data content quality dimension of interactivity of the inquirer. Despite the efforts to mitigate the limitation, the authors acknowledge that secondary data often do not have explicit information about the primary data collection method.

7   Conclusion

This article provides a guideline for a new secondary qualitative data research methodology that draws on a range of existing methods and adds a procedural structure for a complete analysis from the beginning to end, to help remove ambiguity regarding the process. After establishing the research methodology, the guideline begins at formulation of research question, travels through data collection, data quality assessment, thematic discourse data analysis with reflection and reflexivity, and finishes with the self-assessment of findings. Throughout, the method is applied to the context of interviews of people experiencing or

having experienced forced migration by referring to relevant ethical principles. The outlined methodology can help qualitative research of publicly available, online interviews be conducted with a level of rigor to mitigate the otherwise disadvantageous use of secondary data.

Funding Acknowledgments

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

8   References

Abdulla, R. A., Garrison, B., Salwen, M., & Driscoll, P. (2002). The Credibility of
    Newspapers, Television News and Online News. In *Mass communication and society
    division, association for education in journalism and mass communication annual
    conference* (p. 1-30). Miami Beach, Florida. Retrieved from
    `https://citeseerx.ist.psu.edu/viewdoc/`
    `download?doi=10.1.1.556.4841&rep=rep1&type=pdf`

ASA. (2021). *Ethical guidelines for good research practice.* Association of Social
    Anthropologists of the UK (ASA). London, UK. Retrieved from
    `https://www.theasa.org/ethics/`

Australian Red Cross. (2015). *Australian red cross speaks with its new ambassador,
    [interviewee].* Retrieved from `https://www.redcross.org.au/about/governance/`
    `ambassadors/abdi-aden/interview-with-abdi-aden`

Azoulay, K. (2018). *An interview with a syrian refugee on the refugee crisis and ongoing
    civil war.* Retrieved from
    `https://kobiazoulay.medium.com/an-interview-with-a-syrian-refugee-on`
    `-the-refugee-crisis-and-ongoing-civil-war-b419d2fb8010`

Baldwin, J. R., Pingault, J.-B., Schoeler, T., Sallis, H. M., & Munafò, M. R. (2022, jan).
    Protecting against researcher bias in secondary data analysis: challenges and
    potential solutions. *European Journal of Epidemiology*, *37*(1), 1–10. Retrieved from
    `https://doi.org/10.1007/s10654-021-00839-0`   doi: 10.1007/s10654-021-00839-0

Bijak, J. (2022). *Towards bayesian model-based demography.* Springer International
    Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-83039-7`   doi:
    10.1007/978-3-030-83039-7

Bilecen, B., & Lubbers, M. J. (2021). The networked character of migration and
    transnationalism. *Global Networks*, *21*(4), 837–852. Retrieved from

https://doi.org/10.1111/glob.12317 doi: 10.1111/glob.12317

Bögner, D. (2005). *What prevents refugees and asylum seekers exposed to violence from disclosing trauma?* (Doctoral dissertation, University College London, London, UK). Retrieved from https://discovery.ucl.ac.uk/id/eprint/1445321/

Boud, D., Cohen, R., & Walker, D. (Eds.). (1993). *Using experience for learning.* Buckingham, England: Open University Press.

Bozdag, E. (2013, jun). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. Retrieved from https://doi.org/10.1007/s10676-013-9321-6 doi: 10.1007/s10676-013-9321-6

Brackstone, G. (1999). Managing data quality in a statistical agency. *Survey Methodology*, 139–149.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. Retrieved from https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa doi: 10.1191/1478088706qp063oa

BSA. (2021). *Guidelines on Ethical Research.* https://www.britsoc.co.uk/ethics. British Sociological Association (BSA). (Accessed: 2023-04-03)

Caelli, K., Ray, L., & Mill, J. (2003). 'clear as mud': toward greater clarity in generic qualitative research. *International journal of qualitative methods*, *2*(2), 1–13.

Cambridge Dictionary. (n.d.). *Quotation.* Retrieved from https://dictionary.cambridge.org/dictionary/english/quotation (Retrieved September 27, 2022)

Castles, S. (2012). Methodology and methods: Conceptual issues. *African migration research: Innovative methods and methodologies*, 15–36.

Chatfield, S. (2020, mar). Recommendations for secondary analysis of qualitative data. *The Qualitative Report.* Retrieved from https://doi.org/10.46743/2160-3715/2020.4092 doi:

10.46743/2160-3715/2020.4092

Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019, jan). Open data in
    qualitative research. *International Journal of Qualitative Methods*, *18*,
    160940691882386. Retrieved from `https://doi.org/10.1177/1609406918823863`
    doi: 10.1177/1609406918823863

Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The
    relative influence of traditional and technological factors. *Journal of
    Computer-Mediated Communication*, *17*(2), 171–186. Retrieved from
    `https://doi.org/10.1111/j.1083-6101.2011.01565.x` doi:
    10.1111/j.1083-6101.2011.01565.x

Clark-Kazak, C. (2021). Ethics in forced migration research: Taking stock and potential
    ways forward. *Journal on Migration and Human Security*, *9*(3), 125–138. Retrieved
    from `https://doi.org/10.1177/23315024211034401` doi:
    10.1177/23315024211034401

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design* (4th ed.).
    Thousand Oaks, CA: SAGE Publications.

CUREC. (2020). *Studies involving adult refugees in the united kingdom.* University of
    Oxford, Oxford, UK. Retrieved from `https://researchsupport.admin.ox.ac.uk/`
    `files/ap05curecapprovedprocedureidrecv10pdf`

D'Angelo, A. (2021). The networked refugee: The role of transnational networks in the
    journeys across the mediterranean. *Global Networks*. Retrieved from
    `https://doi.org/10.1111/glob.12312` doi: 10.1111/glob.12312

Denzin, N. (1997). *Interpretive ethnography: Ethnographic practices for the 21st century.*
    SAGE Publications, Inc. Retrieved from `https://doi.org/10.4135/9781452243672`
    doi: 10.4135/9781452243672

Dupuy, M.-D. (2012). *Framework for the analysis of separation-related incidents in
    aviation* (Unpublished doctoral dissertation). Imperial College London, London, UK.

Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring
    tools. *Frontiers in Big Data*, *5*. Retrieved from
    `https://doi.org/10.3389/fdata.2022.850611` doi: 10.3389/fdata.2022.850611

Erickson, K., Heald, P., Homberg, F., Kretschmer, M., & Mendis, D. (2015). *Copyright
    and the value of the public domain: An empirical assessment.* Newport, Wales:
    Intellectual Property Office.

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis.
    African Journal of Emergency Medicine, *7*(3), 93-99. Retrieved from
    `https://www.sciencedirect.com/science/article/pii/S2211419X17300423`
    doi: https://doi.org/10.1016/j.afjem.2017.08.001

European Commission. (2021). *Communication from the commission to the european
    parliament and the council: Guidance on article 17 of directive 2019/790 on copyright
    in the digital single market.* Online. Retrieved from `https://eur-lex.europa.eu/`
    `legal-content/EN/TXT/?qid=1625142238402&uri=CELEX%3A52021DC0288`
    (COM/2021/288 final - Document 52021DC0288)

Fedyuk, O., & Zentai, V. (2018). The interview in migration studies: A step towards a
    dialogue and knowledge co-production? In *IMISCOE research series* (pp. 171–188).
    Springer International Publishing. Retrieved from
    `https://doi.org/10.1007/978-3-319-76861-8_10` doi:
    10.1007/978-3-319-76861-8_10

Feinstein, A., Pavisian, B., & Storm, H. (2018, mar). Journalists covering the refugee and
    migration crisis are affected by moral injury not PTSD. *JRSM Open*, *9*(3),
    205427041875901. Retrieved from `https://doi.org/10.1177/2054270418759010`
    doi: 10.1177/2054270418759010

Finlay, L. (2002, aug). Negotiating the swamp: the opportunity and challenge of reflexivity
    in research practice. *Qualitative Research*, *2*(2), 209–230. Retrieved from
    `https://doi.org/10.1177/146879410200200205` doi:

10.1177/146879410200200205

Forss, K. S., Mangrio, E., Leijon, M., Grahn, M., & Zdravkovic, S. (2021). Physical
    activity in relation to wellbeing among newly arrived refugees in sweden: A
    quantitative study. *Frontiers in Public Health*, *8*. Retrieved from
    `https://doi.org/10.3389/fpubh.2020.532883` doi: 10.3389/fpubh.2020.532883

Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions.
    Information Processing & Management, *30*(1), 9–19. Retrieved from
    `https://doi.org/10.1016/0306-4573%2894%2990020-5` doi:
    10.1016/0306-4573(94)90020-5

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C. M., & AOIR. (2020). *Internet
    research: Ethical guidelines 3.0.* Association of Internet Researchers (AOIR).
    Retrieved from `https://aoir.org/reports/ethics3.pdf`

Frydenlund, E., & de Kock, C. (2020, 5). Agent-based modeling within forced migration
    research: A review and critique. In D. Douhaibi & N. Szelei (Eds.), *Refugee review -
    special focus: Emerging issues in forced migration - perspectives from research and
    practice* (Vol. 4, p. 53-68). Online: The Emerging Scholars and Practitioners on
    Migration Issues Network. Retrieved from `https://espminetwork.com/
    wp-content/uploads/2020/06/Refugee-Review-IV.pdf`

Gibson, W. J., & Brown, A. (2009). Using documents in research. In *Working with
    qualitative data* (pp. 65–83). SAGE Publications, Ltd. Retrieved from
    `https://methods.sagepub.com/book/working-with-qualitative-data/d59.xml`
    doi: https://dx.doi.org/10.4135/9780857029041.d59

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for
    qualitative research.* New York: Aldine De Guyter.

Glasgow, R. E. (2013, may). What does it mean to be pragmatic? pragmatic methods,
    measures, and models to facilitate research translation. *Health Education &
    Behavior*, *40*(3), 257–265. Retrieved from

`https://doi.org/10.1177/1090198113486805` doi: 10.1177/1090198113486805

Haegemans, T., Snoeck, M., & Lemahieu, W. (2016). Towards a precise definition of data accuracy and a justification for its measure. In *International conference on information quality* (p. 16:1-16:13). Ciudad Real, June.

Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research*, *33*, 33-45. Retrieved from `http://www.ssoar.info/ssoar/handle/document/19143` doi: 10.12759/HSR.33.2008.3.33-45

Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997, aug). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, *7*(3), 408–424. Retrieved from `https://doi.org/10.1177/104973239700700306` doi: 10.1177/104973239700700306

IASFM. (2018). *Code of ethics.* Toronto, ON, Canada. Retrieved from `http://iasfm.org/wp-content/uploads/2021/11/IASFM-Ethics-EN-compressed.pdf`

Jacobs, M., & Maryns, K. (2021, mar). Managing narratives, managing identities: Language and credibility in legal consultations with asylum seekers. *Language in Society*, *51*(3), 375–402. Retrieved from `https://doi.org/10.1017/s0047404521000117` doi: 10.1017/s0047404521000117

Jacobson, A. F., Hamilton, P., & Galloway, J. (1993, aug). Obtaining and evaluating data sets for secondary analysis in nursing research. *Western Journal of Nursing Research*, *15*(4), 483–494. Retrieved from `https://doi.org/10.1177/019394599301500407` doi: 10.1177/019394599301500407

Kinsella, M. O. (2020). Refugee testimony and human rights advocacy: The challenges of interviewing refugees "in the field". In D. Douhaibi & N. Szelei (Eds.), *Refugee review - special focus: Emerging issues in forced migration - perspectives from research and practice* (Vol. 4, p. 21-31). Online: The Emerging Scholars and Practitioners on Migration Issues Network. Retrieved from `https://espminetwork.com/`

`wp-content/uploads/2020/06/Refugee-Review-IV.pdf`

Klabunde, A., & Willekens, F. (2016, feb). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, *32*(1), 73–97. Retrieved from `https://doi.org/10.1007/s10680-015-9362-0` doi: 10.1007/s10680-015-9362-0

Krause, U. (2017). *Researching forced migration: critical reflections on research ethics during fieldwork.* Refugee Studies Centre, Oxford Department of International Development, University of Oxford. Retrieved from `https://www.rsc.ox.ac.uk/publications/researching-forced-migration -critical-reflections-on-research-ethics-during-fieldwork`

Laranjeiro, N., Soydemir, S. N., & Bernardino, J. (2015). A survey on data quality: Classifying poor data. In *2015 ieee 21st pacific rim international symposium on dependable computing (prdc)* (p. 179-188). doi: 10.1109/PRDC.2015.41

Leggo, C. (2011, jan). Narrative inquiry: Attending to the art of discourse. *Language and Literacy*, *10*(1). Retrieved from `https://doi.org/10.20360/g2sg6q` doi: 10.20360/g2sg6q

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Thousand Oaks, CA: SAGE Publications.

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? trustworthiness and authenticity in naturalistic evaluation. *New directions for program evaluation*, *1986*(30), 73–84.

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica: tecnologies de la traducció*, *12*, 455. Retrieved from `https://doi.org/10.5565/rev/tradumatica.77` (Deutsches Forschungszentrum für Künstliche Intelligenz) doi: 10.5565/rev/tradumatica.77

Mangrio, E., Zdravkovic, S., & Carlson, E. (2018). A qualitative study of refugee families' experiences of the escape and travel from syria to sweden. *BMC Research Notes*,

11(1). Retrieved from `https://doi.org/10.1186/s13104-018-3702-1` doi:
10.1186/s13104-018-3702-1

McAlpine, A., Kiss, L., Zimmerman, C., & Chalabi, Z. (2020, aug). Agent-based modeling
for migration and modern slavery research: a systematic review. *Journal of
Computational Social Science*, *4*(1), 243–332. Retrieved from
`https://doi.org/10.1007/s42001-020-00076-7` doi: 10.1007/s42001-020-00076-7

McCreaddie, M., & Payne, S. (2010). Evolving grounded theory methodology: Towards a
discursive approach. *International Journal of Nursing Studies*, *47*(6), 781–793.
Retrieved from `https://doi.org/10.1016/j.ijnurstu.2009.11.006` doi:
10.1016/j.ijnurstu.2009.11.006

Mehrabi, D., Hassan, M., Sham, M., & Shahkat Ali, M. S. (2009). News media credibility
of the internet and television. *European Journal of Social Sciences*, *11*(1), 136-148.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San
Francisco, CA, USA: Jossey-Bass Publishers.

Migration Data Portal. (2022). *Forced migration or displacement.* Retrieved from `https://www.migrationdataportal.org/themes/forced-migration-or-displacement`

Moher, D., Liberati, A., Tetzlaff, J., & and, D. G. A. (2009, July). Preferred reporting
items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS
Medicine*, *6*(7), e1000097. Retrieved from
`https://doi.org/10.1371/journal.pmed.1000097` doi:
10.1371/journal.pmed.1000097

Müller-Funk, L. (2020). Research with refugees in fragile political contexts: How ethical
reflections impact methodological choices. *Journal of Refugee Studies*, *34*(2),
2308–2332. Retrieved from `https://doi.org/10.1093/jrs/feaa013` doi:
10.1093/jrs/feaa013

O'Connor, H., & Goodwin, J. (2010). Utilizing data from a lost sociological project:

experiences, insights, promises. *Qualitative Research*, *10*, 283 - 298.

Official Journal of the European Communities. (2001). *Directive 2001/29/ec of the european parliament and of the council of 22 may 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.* Online. Retrieved from `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/ ?uri=CELEX:32001L0029&from=EN` (L 167/10 - Document 32001L0029)

Official Journal of the European Union. (2019). *Directive (eu) 2019/790 of the european parliament and of the council of 17 april 2019 on copyright and related rights in the digital single market and amending directives 96/9/ec and 2001/29/ec.* Online. Retrieved from `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/ ?uri=CELEX:32019L0790&from=EN` (PE/51/2019/REV/1 - Document 32019L0790)

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71. Retrieved from `https://doi.org/10.1136/bmj.n71` doi: 10.1136/bmj.n71

Parliament of the United Kingdom. (2021). *Copyright, Designs and Patents Act 1988.* Online. Retrieved from `https://assets.publishing.service.gov.uk/ government/uploads/system/uploads/attachment_data/file/957583/ Copyright-designs-and-patents-act-1988.pdf`

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. Retrieved from `https://doi.org/10.1145/505248.506010` doi: 10.1145/505248.506010

RSC. (2007). Ethical guidelines for good research practice. *Refugee Survey Quarterly*, *26*(3), 162–172. Retrieved from `https://doi.org/10.1093/rsq/hdi0250` doi: 10.1093/rsq/hdi0250

Ruggiano, N., & Perry, T. E. (2019). Conducting secondary analysis of qualitative data: Should we, can we, and how? *Qualitative Social Work*, *18*(1), 81-97. Retrieved from `https://doi.org/10.1177/1473325017700701` (PMID: 30906228) doi: 10.1177/1473325017700701

Ryan, L., & Dahinden, J. (2021). Qualitative network analysis for migration studies: Beyond metaphors and epistemological pitfalls. *Global Networks*. Retrieved from `https://doi.org/10.1111/glob.12319` doi: 10.1111/glob.12319

Sanchez, G., Hoxhaj, R., Nardin, S., Geddes, A., Achilli, L., & Kalantaryan, R. S. (2018). *A study of the communication channels used by migrants and asylum seekers in italy, with a particular focus on online and social media.* European Commission and European University Institute and Directorate-General for Migration and Home Affairs and Robert Schuman Centre for Advanced Studies. Retrieved from `https://op.europa.eu/en/publication-detail/-/publication/` `2af2973d-48fc-11e8-be1d-01aa75ed71a1` doi: doi/10.2837/77854

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340. Retrieved from `https://doi.org/10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g` doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g

Savin-Baden, M., & Howell, C. (2012). *Qualitative research.* London, England: Routledge.

Sindin, X. P. (2017). Secondary data. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods* (p. 1578-1579). Thousand Oaks: SAGE Publications, Inc. Retrieved from `https://methods.sagepub.com/reference/` `the-sage-encyclopedia-of-communication-research-methods/i13206.xml` doi: https://dx.doi.org/10.4135/9781483381411.n557

State of New South Wales (Department of Education). (2020). *DVD: Who is a Refugee.* Retrieved from `https://www.roads-to-refuge.com.au/resources/transcripts/` `transcript-who-is-a-refugee.html`

Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data. *JAMA*, *313*(16), 1657. Retrieved from `https://doi.org/10.1001/jama.2015.3656` doi: 10.1001/jama.2015.3656

Szabo, V., & Strang, V. R. (1997, dec). Secondary analysis of qualitative data. *Advances in Nursing Science*, *20*(2), 66–74. Retrieved from `https://doi.org/10.1097/00012272-199712000-00008` doi: 10.1097/00012272-199712000-00008

Taylor, G. W., & Ussher, J. M. (2001). Making sense of s&m: A discourse analytic account. *Sexualities*, *4*(3), 293-314. Retrieved from `https://doi.org/10.1177/136346001004003002` doi: 10.1177/136346001004003002

Technical Committee ISO/TC37. (2015). *Translation services — Requirements for translation services* (Standard). Geneva, CH: International Organization for Standardization.

The Intellectual Property Office. (2014). *Exceptions to copyright: Research.* Online. The Intellectual Property Office, Concept House, Cardiff Road, Newport, NP10 8QQ, United Kingdom. Retrieved from `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf`

Thorne, S. (1994). Secondary analysis in qualitative research: Issues and implications. *Critical issues in qualitative research methods*, *1*, 263–279.

Thorne, S. (1998). Ethical and representational issues in qualitative secondary analysis. *Qualitative Health Research*, *8*(4), 547–555. Retrieved from `https://doi.org/10.1177/104973239800800408` doi: 10.1177/104973239800800408

Tripathy, J. P. (2013, 12). Secondary data analysis: Ethical issues and challenges. *Iranian*

Journal of Public Health, *42*(12), 1478–1479. Retrieved from

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/

United Nations General Assembly. (1951). Convention Relating to the Status of Refugees.
United Nations, Treaty Series, *189*, 137-189. (1951 Refugee Convention)

United States Copyright Office, Library of Congress. (2021). *Copyright law of the united
states and related laws contained in title 17 of the united states code.* Online.
Retrieved from https://www.copyright.gov/title17/title17.pdf (Circular 92 -
05/2021)

West, M. (1993). *Developing high quality data models* (Tech. Rep. No. 3). London, UK:
European Process Industries STEP Technical Liaison Executive.

Yin, R. K. (2018). *Case study research and applications* (6th ed.). Thousand Oaks, CA:
SAGE Publications.

Table 1

Summary of data quality dimensions in alphabetical order

Cat*	Dimension	Description	References
Data Context	Accessibility	Only data that is publicly available are to be used. This means that the data can be retrieved without a prohibition of access by technological protection measures such as username and password.	Dupuy (2012); Hinds et al. (1997); Laranjeiro et al. (2015); Pipino et al. (2002)
	Appropriate amount of data	There is no set number of required word count, but there needs to be enough sentences from the interviewee alone within the data–without including the words of the interviewers–for the data to be processed and understood without needing more context.	Laranjeiro et al. (2015); Pipino et al. (2002)
	Biasness or objectivity	The interviewer cannot be leading questions to have their own story told. The quotation or interviews cannot be filled with the publishing representatives' words. Particularly, the possibility of selective reporting needs to be assessed based on questions and surrounding information.	Mehrabi et al. (2009); Patton (2002)
	Conformance or confirmity	Only a specific format of data is to be included in the analysis for general consistency of format (but not necessarily consistency in content nor questions). The data need to be an interview or quotation. The interviewees need to be from the target population.	Ehrlinger & Wöß (2022)
	Credibility or integrity of source	The credibility of the publisher needs to be assessed through a background check of the organization, especially on whether the organization works with integrity. This is to assess if the organization has any hidden agenda that could influence the publication of the interview.	Ehrlinger & Wöß (2022); Laranjeiro et al. (2015); Pipino et al. (2002)
	Duplication or Duplicates	Any duplicates in the datasets are to be removed.	Ehrlinger & Wöß (2022)
	(Hyper-) emotivity	Any data where the interviewer or the surrounding information uses hyper-emotional language is to be excluded, as this influences the interview setting and questions the overall credibility, even if the interviewer may have been using emotions to show support or sympathy.	Added for this study
	Interactivity of the inquirer	Any data where the interviewer or the publisher interferes, interjects, or interrupts the interviewee are excluded. The narrative needs to be from the interviewee.	Added for this study
	Multimediality	Interviews in both text and video formats are included. Use of photos of the interviewees and hyperlinks to other references help provide credibility of the interviews. However, within videos if there is an influential background music, the entire interview is to be excluded from the dataset to remove bias.	Chung et al. (2012)

* Cat = Category

Table 1

Summary of data quality dimensions in alphabetical order (continued)

Cat*	Dimension	Description	References
Data Context *(continued)*	Relevance or relevancy	The data and quotations need to relate to the research question.	Dupuy (2012); Hinds et al. (1997); Laranjeiro et al. (2015); Pipino et al. (2002)
	Trust or trustworthiness	Although related to credibility (of the publisher and interviewer), this dimension examines the questions from the interviewer and whether there is a hidden agenda or goal-oriented strategy beyond allowing the interviewee to express their own narrative.	Chung et al. (2012); Mehrabi et al. (2009); Patton (2002)
	Uniqueness or Value-added	The benefit of using secondary analysis is diversity. Therefore, variation sampling method can help obtain interviews from a variety of demographics. This is an inclusion and not an exclusion criteria.	Ehrlinger & Wöß (2022); Laranjeiro et al. (2015); Pipino et al. (2002)
Data Content	Appropriate amount of relevant data	There needs to be enough data to get to know the interviewee and yet also enough data about the target sample's perspective on the secondary research question.	Adapted from Dupuy (2012); Laranjeiro et al. (2015); Pipino et al. (2002)
	Authenticity	This dimension questions whether the words of interviewees have been altered for the interest of the publisher in the form of promotion of publisher within their answer. Different to the unused dimension of consistency, where the same emotional approach or viewpoint throughout the interview is expected, this dimension evaluates the likelihood of a post-interview edit by the publisher. This is not meant for questioning the interviewee's words, as they represent the perspective, experience, and answer of the interviewee.	Patton (2002)
	Clarity, interpretability, understandability	This dimension evaluates whether the data can be understood by the authors. Although disconnected words can be a result of memory lapse, if the sentence is difficult for the authors to interpret, this part of the data needs to be excluded from the analysis to avoid misinterpretation.	Dupuy (2012); Laranjeiro et al. (2015); Mehrabi et al. (2009); Pipino et al. (2002)
	(Self-) representation	The data needs to be about the interviewee and their perception of their experience. Any data where the interviewee talks about someone else's lives or views they have not directly witnessed have been further assessed for relevancy.	Added for this study

* Cat = Category

Table 2

Proposed phases of secondary qualitative research process

| Phase of secondary data analysis	Description of the process	
1. Formulation of research questions	The research question needs to be defined prior to the analysis to mitigate researcher bias (Baldwin et al., 2022).	
2. Data collection	Retrieval of data using web search of non-personalized results for online, publicly accessible secondary qualitative data in interview format, as shown in section 5.2 Step 2. Data Collection. Using new, unfamiliar data without prior knowledge is important to minimize the risk of bias (Baldwin et al., 2022).	
3. Transcription, if applicable	Although much of the data may already be in written and transcribed format, any video and/or audio data needs to be transcribed verbatim.	
4a. Data filtering - Data context quality assessment	This is the first step of filtering the data based on the quality of the data source and interviewer questions and verbal interaction with the participant as well as reporting technique, as detailed in section 5.4 Step 4. Data Filtering.	
4b. Data filtering - Data content quality assessment	This is the second step of filtering the data based on the content quality and its appropriateness for the research question, as detailed in section 5.4 Step 4. Data Filtering.	
5. Recognition of researcher positionality	"Prospective reflection" (Boud et al., 1993) is encouraged with a positionality statement to explicitly explore the researcher's personal stance on the overall subject, acknowledge its impact on the analysis, and investigate possible mitigation.	
6a. Data analysis - Familiarization with data	This is different to having prior knowledge of data prior to analysis. Familiarization, on the other hand, is necessary to gain initial insights of the new data (Braun & Clarke, 2006). Although the researchers may have developed a basic idea during the data filtering stages, data needs to be read (again) without distraction.	Continuous reflection and reflexivity (Finlay, 2002)
6b. Data analysis - Content analysis and categorization	Data needs to be read with intention of content analysis (Erlingsson & Brysiewicz, 2017) to develop categorization and metadata of each text.	
6c. Data analysis - Open coding	Initial codes need to be generated (Braun & Clarke, 2006) using open coding method based on discursive grounded theory (McCreaddie & Payne, 2010) and narrative research approach (Creswell & Poth, 2017) to detect patterns specific to individual experiences.	
6d. Data analysis - Axial coding for initial themes	The codes can be laid out in multiple formats to look for categories and patterns (Braun & Clarke, 2006) using axial coding methodology (McCreaddie & Payne, 2010).	
6e. Data analysis - Developing final themes	Various categorizations need to be trialed to understand the best overall themes, particularly based on data availability and research questions. The final themes need to be defined in such way that they can help answer the research questions.	
6f. Data analysis - Answering the research questions	The data then needs to be regrouped to each theme and a selection of the extracts need to be presented to support the analysis and answer the research questions (Braun & Clarke, 2006).	
7. Assessment of trustworthiness of the findings	In addition to self-assessment statements of each step, using the criteria of credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1985), the trustworthiness of the data analysis and findings need to be assessed, as detailed in subsection 5.6 Step 6. Data Analysis.	
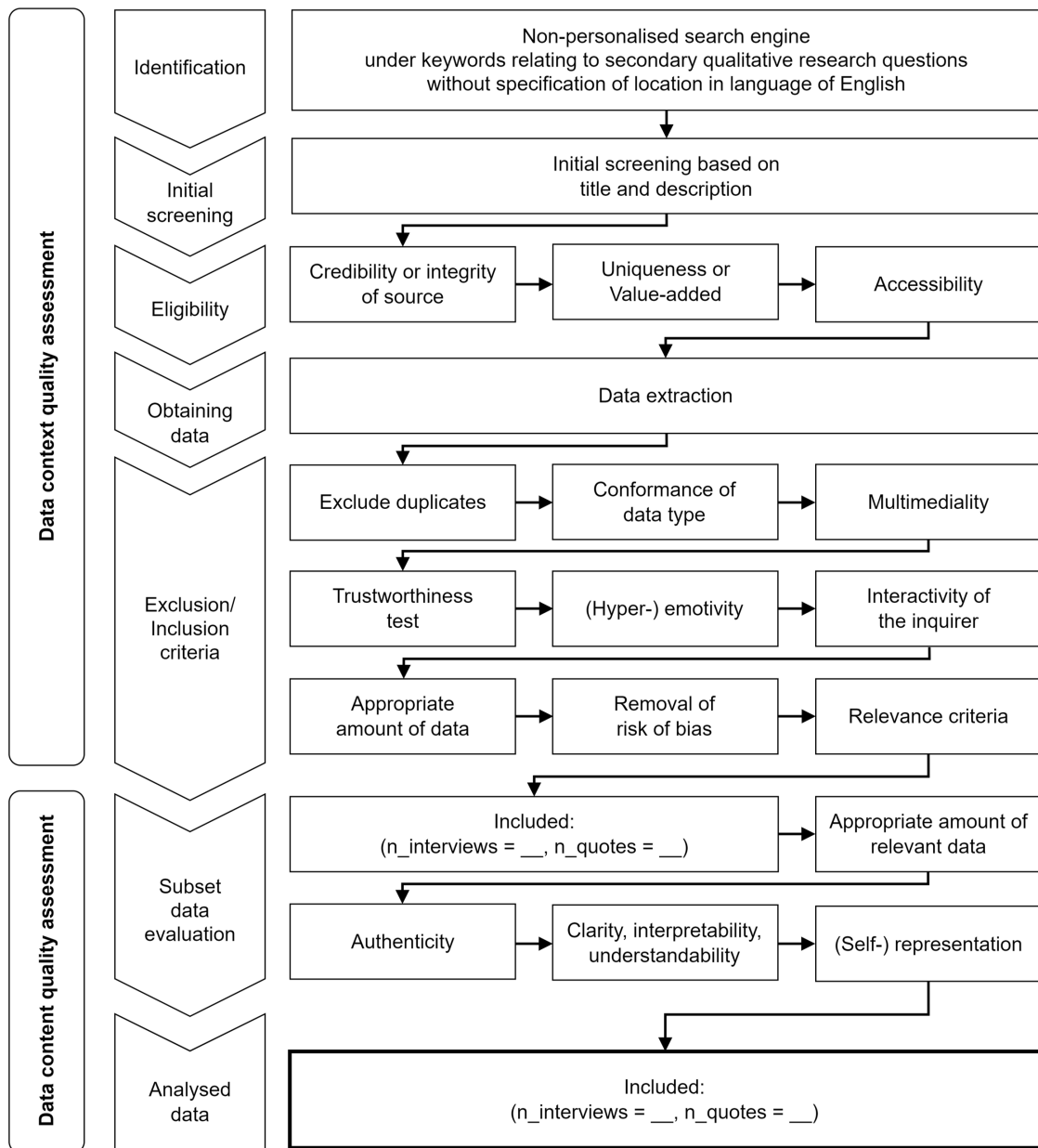
Figure 1. Data filtering procedure based on PRISMA