Developing codes from the interview: reading versus listening

Titan Ligita, Karen Francis, Kristin Wicking et al

Abstract

Background Interviewing is a vital and common method of collecting data in qualitative research. The interview is usually recorded and a written transcription is created from the recording. The transcription document is then analysed by reading and re-reading to fracture the data and develop initial codes, as in grounded theory methodology. However, this method has disadvantages.

Aim To report on how the authors used the process of generating initial codes during their analysis in a research study.

Discussion The authors compare the rigour and efficiency of generating initial codes from reading written transcripts with generating initial codes from listening to recordings. The most notable difference between the two methods is the length of time needed to transcribe the recording before coding can start. The authors discuss the lessons they learned from their pragmatic decision to expedite initial coding by listening to rather than reading the interview data.

Conclusion Grounded theory requires concurrent data generation and analysis. Audio analysis is efficient in developing initial codes from interview recordings.

Implications for practice Nurse researchers can use the audio method of analysing interview data

Author details

Titan Ligita, lecturer, School of Nursing, Faculty of Medicine, Universitas Tanjungpura, Pontianak, Indonesia; Karen Francis, professor of nursing, midwifery and nutrition, College of Healthcare Sciences, University of Tasmania, Hobart Tasmania, Australia; Kristin Wicking, senior lecturer, College of Healthcare Sciences, James Cook University, Townsville, Queensland, Australia; Nichole Harvey, associate professor, College of Medicine and Dentistry, Townsville, Queensland, Australia; Intansari Nurjannah, associate professor, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Introduction

A qualitative research study is highly informative as it provides rich data. When conceptualising a study, researchers need to consider carefully which methods of collecting data they will use and whether these methods will answer their research questions. Researchers must also decide

which methods they will use to analyse the data they collect. Their analysis will assist them to generalise the data theoretically and analytically rather than statistically (Eakin and Gladstone 2020).

This article details the lessons learned while conducting a grounded theory study. These lessons are explained to other researchers for their consideration to possibly adopt in their own work.

Background

The methods used to collect data in a study must be aligned with the research questions, to ensure it is of a high quality. Several types of data can be generated in qualitative studies, including interviews, observations, document reviews, field notes, picture and videos (Corbin and Strauss 2008). Most data generated are textual – either spoken or written.

Qualitative researchers investigate not only what has happened but also how it has occurred (Teti et al 2020). By collecting qualitative data, researchers can discover different views, thoughts and meanings from participants in their studies (Teti et al 2020). Interviews are the most commonly chosen method of collecting qualitative data, as it is easier to obtain insights into a phenomenon by questioning the participants than it is by observation (Holloway and Galvin 2017).

Transparency is also crucial in qualitative research, to show that the findings are trustworthy. Transparency generates trust in the analysis of the data, especially when fracturing the data processed from the spoken word into a written transcript.

The raw data recorded from participants or events cannot explain the meaning of the phenomena being studied; instead, analysis reveals the meaning inherent in the data (Pope et al 2000). The analytical methods used in qualitative research are intended to assist researchers in dealing with complex data by reducing them to a smaller quantity while still maintaining the quality of the concepts (Corbin and Strauss 2008).

The complexity of qualitative data means that researchers need to use a reliable, systematic method of analysis to explain the phenomena being studied. The commonly used inductive approach in which original raw data are the source for the analysis is time-intensive because it requires considerable effort (Burnard et al 2008). However, the inductive approach is appropriate for studying a phenomenon about which there is limited knowledge (Burnard et al 2008). Therefore, opportunities to increase the efficiency of the inductive process without sacrificing its rigour are worthy of consideration.

There are different types of inductive approaches for analysing qualitative data. Thematic analysis and content analysis are the approaches most commonly used in nursing research (Vaismoradi et al 2013). Constant comparative analysis is an essential method in grounded theory, which is a popular methodology intended to generate theory from the actual data (Birks and Mills 2015).

Despite the variety of approaches used in qualitative data analysis, they have one commonality – the data need to be fractured into smaller units. These units are generally referred to as 'themes', although grounded theory calls them 'codes'. Themes report how the primary raw data have been compressed (Morgan and Nica 2020). Researchers are less likely to construct themes or categories directly if they do not initially break down the data into smaller pieces.

To develop sensible data, the researcher needs to 'open' the codes (Birks and Mills 2015). This involves identifying 'important words or groups of words in the data' and attaching a suitable label (Birks and Mills 2015). Open coding is an explorative process that leads to the identification of a concept from the collected data (Corbin and Strauss 2015). The essential role of coding in grounded theory is to create a bridge from the collected data to the emerging theory (Charmaz 2014). Coding assists researchers in identifying what is

in the data and elucidating participants' implied processes, actions and views. Researchers can closely examine and analyse the data through coding (Charmaz 2014). A code can be 'a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data' (Saldaña 2013). Codes can therefore be generated from spoken and written texts.

Researchers can develop their own methods of analysing data. However, they must be flexible and responsive to the data, the study's aim and the time available for analysis (Corbin and Strauss 2015).

Some qualitative research papers have discussed the management of interview data (Wellard and McKenna 2001, Halcomb and Davidson 2006) but despite the complexity and volume of qualitative data, there is limited information about what method of analysing data is best. The method researchers choose to analyse the data will have practical implications for time efficiency, especially if they are undertaking concurrent analysis in the field as they will need to pay for daily expenses such as accommodation, meals and transport.

Analysis is typically conducted through transcription (Halcomb and Davidson 2006). Complete verbatim transcription includes the reproduction of the words spoken by participants as writing, as well as of the emotional context of the language spoken, by using standardised syntax in the written texts and recording non-verbal vocalisations such as sighs, laughter or sobs (Poland 1995).

Researchers commonly use transcriptions as the first step when analysing the data. However, verbatim transcription may take time and there is no evidence it is the most appropriate technique to use to develop initial codes.

Method
This paper is part of a more extensive study which investigated how people with diabetes in Indonesia learn about their disease (Ligita et al 2019a). The study

was conducted in a province of Indonesia, with data collected between April 2016 and July 2017. Grounded theory was chosen to generate a theory about health education for people living with diabetes, as the research question asked about the process by which people acted and performed interactions (Birks and Mills 2015). The essential methods in grounded theory – such as selecting the participants and performing concurrent data collection and analysis – modified by Birks and Mills (2015) were followed to ensure the study was rigorous (Ligita et al 2019a).

Twenty six participants were recruited to the study using purposeful sampling during the initial phase of data collection and theoretical sampling to generate further data (Ligita et al 2020). The participants were people with diabetes, healthcare professionals, family members, lay health workers, hospital health promotion staff and an exercise instructor. The primary source of data was audio-recorded, semi-structured interviews. In the first two phases, the lead researcher (TL) conducted the interviews in person or by phone during two field trips in Indonesia; she conducted the interviews in the third phase remotely from Australia by phone.

Two different approaches to analyse the data collected during the field trips were used.

TL conducted and audio-recorded seven interviews during the first field trip. The average length of the interview was approximately 28 minutes. The first of these interviews was transcribed before being analysed. However, this was time-consuming and ineffectual, so the process was modified for the other six interviews.

TL interviewed 17 people during the second field trip. The average length of the interview was approximately 21 minutes. These interviews were initially analysed by listening to each recording and establishing preliminary codes concurrently. The recordings were later transcribed when TL returned to Australia, to support the refinement of the overall analysis process.

Key points
- Reading the transcript and listing to audio recordings assist nurse researchers in analysing data in a grounded theory
- Developing initial codes by listening to a recording may maintain study rigour through immersion in the data
- Listening to interview recordings may take less time

The third phase was intended to verify the data for existing and new participants. Data about participants' experiences, views and insights were summarised and then shared with the other participants to explain and validate the data (Rix et al 2020). The field trip parameters did not constrain the analysis of these four interviews, and so is not the focus of this paper.

Translation

Two languages were used at various times during the study – English and the official Indonesian language spoken by the participants (Ligita et al 2019b). The informed consent form and a participant information letter needed to be translated from English into Indonesian in preparation for data collection. Indonesian was used in the interviews. Both English and Indonesian were used during analysis, since our team consisted of both Indonesian- and English-speaking researchers.

Before achieving theory integration, the study's results were distributed in the third phase to new and existing participants as a storyline and a concept map to confirm the analysed data (Ligita et al 2019b). This process was intended to ensure consistency in the meaning of the storyline.

Both languages were used presenting the final results, which included the translation and confirmation of the analysed data by the bilingual advisor (IN) and TL, who both spoke English and Indonesian. We also published one article in Indonesian to demonstrate the study's results (Ligita et al 2019c).

Results

We used two different methods to analyse the data: conventional data analysis and audio-recording analysis. Conventional data analysis involved reading the transcript of an interview; in contrast, audio-recording analysis involved listening to the recording of an interview. We compared the methods to determine whether reading a transcript of an interview or listening to a recording of the interview was more

effective and efficient in developing codes. Both processes are verbatim and aim to develop initial codes.

The first field trip

Two of us (TL, IN) were involved in the initial phases of data analysis. TL interviewed all the participants during the first field trip. She was therefore closer to the data, so transcribed the interviews verbatim herself. It took approximately half an hour to transcribe every 10 minutes of an interview into an excerpt. TL then developed codes by thoroughly reading every line of the transcripts (Charmaz 2006).

TL and IN met during the final week of the field trip. IN read the first transcript to develop codes from it; TL and IN then compared their codes for the interview (Figure 1). This method was time-consuming as the transcript was decontextualised for IN, who had not been present during the interview.

This experience proved cumbersome, so the method was changed for subsequent transcripts – IN listened to TL read the transcripts aloud. As TL had conducted the interviews, she could almost role-play, reading the transcript with the tonal inflections and emotions she recalled the participant using during the interview. It flowed more naturally and IN understood it more easily, with pauses and repetitive words being given the flavours of hesitancy or emphasis present at the time of the interview. Such repetitions and pauses can make a written transcript cumbersome to read and interpret, but it 'came alive' when read out by TL, the additional detail and precision becoming an aid instead of a barrier to IN's understanding of the meanings of the participant's words, pauses, signs, inflections and so on. After all the transcripts had been coded, TL and IN discussed the developed codes and reached a consensus about them.

This strategy was faster and easier to understand than reading the transcript to IN. However, it was still time-consuming as

transcribing the interviews in preparation for the meetings took TL longer. The total time needed to develop the codes was approximately 31.5 hours: about 3.5 hours for the interviews; approximately 10.5 hours for transcribe the recordings; approximately seven hours for TL to develop codes; and a further 10.5 hours for consensus coding.

Table 1 shows examples of the initial codes developed using the two methods.

The second field trip
The analysis for the second field trip was intended to fracture the data to develop initial codes from the interviews. IN and TL met and listened to each of the 17 audio recordings together. Both authors independently produced notes while listening to the recordings and during a structured and recurring pause every five minutes. After listening to all the interviews, IN and TL compared notes about the codes they had developed. The two authors then re-listened to all the interviews together.

This activity was efficient and effective as the codes were developed closer to the point the data were collected and in a shorter time frame – approximately 24 hours for 17 interviews as opposed to 31.5 hours for seven interviews in the first phase.

Following TL's return to Australia, the audio-recordings from the second field trip were transcribed for subsequent uploading into a program for data management, such as NVivo.

Discussion
Reading the transcripts
Advantages
One of the strengths of reading the transcripts was the ability to capture more details, both important and perhaps less important to the research question.
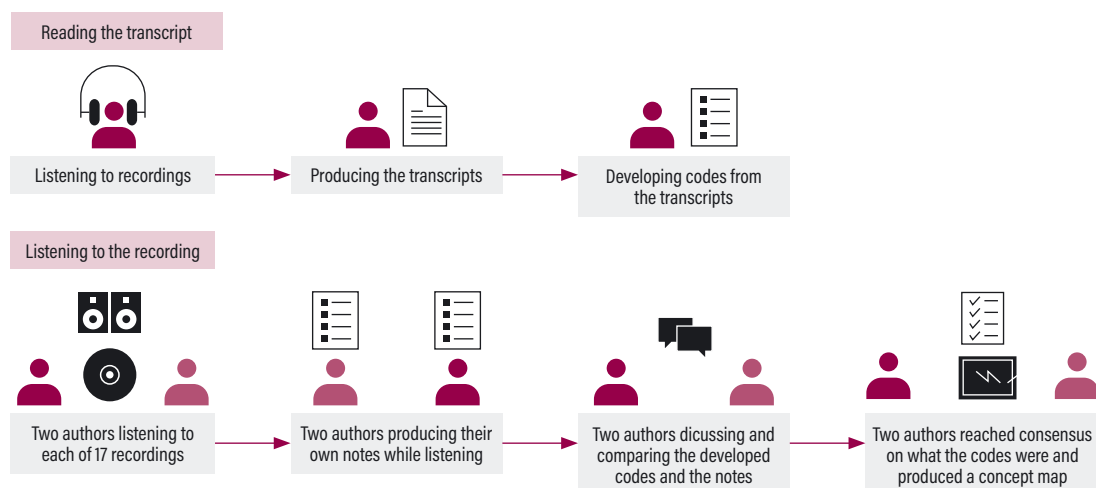
The researchers could immerse themselves in the data and know the data intimately through the process of transcription. By transcribing interviews themselves, researcher can be closer to the data; this aids analysis and the establishing of an audit trail for the process (Halcomb and Davidson 2006).

In analysing the transcripts, memos and notes were written. This promotes the use of reflexivity in a grounded theory study.

Disadvantages
The researchers needed to revisit the original recordings to ensure their agreement about the codes and categories generated came from the participants' voices, not their own beliefs. This step required much more time. It also required more time to first transcribe the interviews

Figure 1. Developing codes from interviews by reading the transcripts and listening to the recordings

Reading the transcript

Listening to recordings → Producing the transcripts → Developing codes from the transcripts

Listening to the recording

Two authors listening to each of 17 recordings → Two authors producing their own notes while listening → Two authors dicussing and comparing the developed codes and the notes → Two authors reached consensus on what the codes were and produced a concept map

into text and then produce codes, especially when dealing with a significant number of interviews. This may inhibit analysis in a grounded theory study, which requires a comparison of data sets during the constrained time of a field trip.

Transcription requires time, physical energy and human resources; it also has a cost burden, especially when other transcribers are employed (Halcomb and Davidson 2006).

Listening to recordings
Advantages and disadvantages
This method is time-efficient and maintains the rigour expected in qualitative research. Time efficiency is assured, as recordings can be transcribed after the field trip has concluded, without the same constraints on time and resources. One study discussed in Halcomb and Davidson (2006) did not use verbatim transcription at all – the researchers analysed their interview data by listening to the recordings several times. Halcomb and Davidson (2006) offered reflexive strategies that included taking notes, reflective journaling, revising field notes, analysis of preliminary and secondary content, and thematic review to manage interview data.

Halcomb and Davidson (2006) also contended that researchers can explore innovative methods for managing and analysing interview data, as long as the methods are reliable and agree with the underpinning methodology.

Rosenblum et al (2004) suggested that coding from a recording – what Parameswaran et al (2020) called 'live coding' – is cost-effective and produces richer data than coding from a transcript.

It is also quicker as much more time is needed to listen to and transcribe the interview than to only listen to the recording. To fracture the data and develop initial codes, listening to the interview recordings was at least two times faster in this present study than creating and then reading the transcripts.

Listening to interviews assists researchers to immerse themselves in conversations with participants. It enables them to hear variations in tone that may provide insights and language nuances (Halcomb and Davidson 2006), as well as emphasis of particular words, phrases or ideas, which can inform data analysis. Listening directly to the participant's voice in a recording may increase the researcher's sensitivity to the original interview data, as it highlights the participant's emphasis and tone (Graneheim and Lundman 2004). Researchers' cultural competency may enable them to understand the participant's behaviour (Gu 2020). Graneheim and Lundman (2004) pointed out that the researcher may notice which parts of a participant's experiences make them feel more important as they may emphasise those parts with a particular tone of voice, inflection or non-verbal vocalisations.

The researcher can write notes immediately after listening to an interview. As with a transcript, these notes will be a transition from the spoken word to the written word, but the notes will also have undergone an additional level of sifting and analysis, as the researcher listens deeply to the underlying meaning and

Table 1. Examples of initial codes generated using the two methods of analysis

Participants' voices	Codes (Transcription)	Codes (Audio-recording)
The dry the wet [diabetes]... I don't know about the dry the wet [diabetes]	Patient unfamiliar with the terms used	Not knowing the term used
If not having injection, where to go doc? I cannot have injection	Patient response: Bargaining	Bargaining for other options
After that, I moved to a public health centre consuming glibenclamide for 10 years, but nothing changes, then I move to doctor B, the internist	No changes, making the patient decide to move to another doctor	Seeking another physician since no progress
There were blacks on all the injection marks... 'The needle must be changed, mam,' he [the doctor] said... 'After three days, change the needle,' he said	Giving instruction related to injection based on the patient's report of the side effect	The patient learns that the bruise occurred because of the unchanged needles

then notes and records its essence. This helps them to develop their theoretical sensitivity and be alert to the same patterns if they recur in other participants' interviews. They can also code better as they literally hear the participants' voices, which convey their thoughts, feelings and ideas more authentically and directly than a transcript can. Researchers therefore need to have training in interviewing to obtain better quality data and be more responsive (Rutakumwa et al 2020).

In this study, two of the authors listened to the audio recordings simultaneously, which added value to the analysis by ensuring the codes developed were thoroughly grounded in the data. Rechecking the tentative generated codes with the recording is easier when coding every five minutes than when listening to and simultaneously coding a whole interview. However, this technique may not be appropriate for generating codes in detail because the author needs to read the transcripts thoroughly.

Part of the essential method of grounded theory is that when developing initial codes, researchers must concurrently collect and analyse data. Therefore, once they finish collecting one set of data, they need to analyse it. By listening to the recording of one interview, they can develop the initial codes promptly and decide which further data they need in the upcoming interviews. This is called theoretical sampling (Ligita et al 2020).

After returning from the field trips, TL transcribed each of the 17 interviews and checked these transcripts to ensure she had not missed any important points or other initial codes during the field trips. Revisiting these raw data also assisted in achieving consensus over codes by avoiding inaccuracies about the meaning of data. Therefore, early and regular communication between researchers is needed (Giesen and Roeser 2020). When reaching a consensus, the researchers also determined that the translation into English of Indonesian coding labels

encapsulated their essence. The transcripts were then used for intermediate and advanced coding, as per the typical qualitative data analysis processes seen in qualitative studies, including grounded theory.

Conclusion

This novel paper compared two methods of analysing data to develop initial codes in a grounded theory study: conventional written data analysis and audio-recording analysis. This comparison will be helpful for nurse researchers interacting with the original raw interview recordings when undertaking qualitative data analysis, including grounded theory analysis.

Researchers may choose to either listen to the recording or read the written transcript, to fracture data and develop initial codes. The option they choose will influence how efficiently they can proceed with the initial coding. In our study, listening to the interview recordings was at least twice as fast at developing the initial codes as reading the written transcripts. Time efficiency is crucial when undertaking field work, with its concomitant resource usage.

Listening to the participants' voices also strengthened the analysis in two ways. First, the recordings preserved all elements of the data such as inflections, tone, emphasis and non-verbal vocalisations, enriching this data. Second, having both researchers listening concurrently to the interviews meant each could appreciate the full spectrum of the audible data. This assisted them in reaching a consensus about what the initial codes should be.

TL frequently visited the raw data to minimise the risk of decontextualised codes and discussed any developed codes with the rest of the team to achieve the most proximate meaning to participants' voices. This process also enabled translation of the codes from Indonesian to English, with TL and IN confirming the translations captured the essence of the Indonesian coding labels.

References

Birks M, Mills J (2015) Grounded Theory: A Practical Guide. Second edition. Sage Publications, Thousand Oaks CA.

Burnard P, Gill P, Stewart K et al (2008) Analysing and presenting qualitative data. British Dental Journal. 204, 429-432. doi: 10.1038/sj.bdj.2008.292

Charmaz K (2006) Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Sage Publications, London.

Charmaz K (2014) Constructing Grounded Theory. Second edition. Sage Publications, London.

Corbin JM, Strauss AL (2008) Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Third edition. Sage Publications, Los Angeles CA.

Corbin JM, Strauss AL (2015) Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Fourth edition. Sage Publications, Los Angeles CA.

Eakin JM, Gladstone B (2020) 'Value-adding' analysis: doing more with qualitative data. International Journal of Qualitative Methods. 19, 1-13. doi: 10.1177/1609406920949333

Giesen L, Roeser A (2020) Structuring a team-based approach to coding qualitative data. International Journal of Qualitative Methods. 19, 1-7. doi: 10.1177/1609406920968700

Graneheim UH, Lundman B (2004) Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Education Today. 24, 2, 105-112. doi: 10.1016/j.nedt.2003.10.001

Gu C-J (2020) Qualitative interviewing in ethnic-Chinese contexts: reflections from researching Taiwanese immigrants in the United States. International Journal of Qualitative Methods. 19, 1-11. doi: 10.1177/1609406920910319

Halcomb EJ, Davidson PM (2006) Is verbatim transcription of interview data always necessary? Applied Nursing Research. 19, 1, 38-42. doi: 10.1016/j.apnr.2005.06.001

Holloway I, Galvin K (2017) Qualitative Research in Nursing and Healthcare. Fourth edition. Wiley Blackwell, Oxford.

Ligita T, Wicking K, Francis K et al (2019a) How people living with diabetes in Indonesia learn about their disease: a grounded theory study. PLoS ONE. 14, 2, e0212019. doi: 10.1371/journal.pone.0212019

Ligita T, Francis K, Wicking K et al (2019b) Using storylines for bilingual dissemination of a grounded theory. Nurse Researcher. 27, 4, 13-18. doi: 10.7748/nr.2019.e1652

Ligita T, Nurjannah I, Harvey N et al (2019c) Sebuah storyline mengenai proses bagaimana penderita diabetes di indonesia mempelajari mengenai penyakitnya Indonesian Journal of Nursing Practices. 3, 1, 9-21 doi: 10.18196/ijnp.3187

Ligita T, Harvey N, Wicking K et al (2020) A practical example of using theoretical sampling throughout a grounded theory study: a methodological paper. Qualitative Research Journal. 20, 1, 116-126. doi: 10.1108/QRJ-07-2019-0059

Morgan DL, Nica A (2020) Iterative thematic inquiry: a new method for analyzing qualitative data. International Journal of Qualitative Methods. 19, 1-11. doi: 10.1177/1609406920955511

Parameswaran UD, Ozawa-Kirk JL, Latendresse G (2020) To live (code) or to not: a new method for coding in qualitative research. Qualitative Social Work. 19, 4, 630-644. doi: 10.1177/1473325019840394

Poland BD (1995) Transcription quality as an aspect of rigor in qualitative research. Qualitative Inquiry. 1, 3, 290-310. doi: 10.1177/107780049500100302

Pope C, Ziebland S, Mays N (2000) Qualitative research in health care: analysing qualitative data. BMJ. 320, 7227, 114-116. doi: 10.1136/bmj.320.7227.114

Rix J, Carrizosa HG, Sheehy K et al (2020) Taking risks to enable participatory data analysis and dissemination: a research note. Qualitative Research. 22, 1, 143-153. doi: 10.1177/1468794120965356

Rosenblum KL, Zeanah C, McDonough S et al (2004) Video-taped coding of working model of the child interviews: a viable and useful alternative to verbatim transcripts? Infant Behavior and Development. 27, 4, 544-549. doi: 10.1016/j.infbeh.2004.04.001

Rutakumwa R, Mugisha JO, Bernays S et al (2020) Conducting in-depth interviews with and without voice recorders: a comparative analysis. Qualitative Research. 20, 5, 565-581. doi: 10.1177/1468794119884806

Saldaña J (2013) The Coding Manual for Qualitative Researchers. Second edition. Sage Publications, London.

Teti M, Schatz E, Liebenberg L (2020) Methods in the time of COVID-19: the vital role of qualitative inquiries. International Journal of Qualitative Methods. 19, 1-5. doi: 10.1177/1609406920920962

Vaismoradi M, Turunen H, Bondas T (2013) Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nursing & Health Sciences. 15, 3, 398-405. doi: 10.1111/nhs.12048

Wellard S, McKenna L (2001) Turning tapes into text: issues surrounding the transcription of interviews. Contemporary Nurse. 11, 2-3, 180-186. doi:10.5172/conu.11.2-3.180