The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections

F C Zamawe

1. Durham University - Anthropology, Durham , Durham DH1 3LE United Kingdom

Correspondance: fzamawe@gmail.com

Abstract

For a long time, electronic data analysis has been associated with quantitative methods. However, Computer Assisted Qualitative Data Analysis Software (CAQDAS) are increasingly being developed. Although the CAQDAS has been there for decades, very few qualitative health researchers report using it. This may be due to the difficulties that one has to go through to master the software and the misconceptions that are associated with using CAQDAS. While the issue of mastering CAQDAS has received ample attention, little has been done to address the misconceptions associated with CAQDAS. In this paper, the author reflects on his experience of interacting with one of the popular CAQDAS (NVivo) in order to provide evidence-based implications of using the software. The key message is that unlike statistical software, the main function of CAQDAS is not to analyse data but rather to aid the analysis process, which the researcher must always remain in control of. In other words, researchers must equally know that no software can analyse qualitative data. CAQDAS are basically data management packages, which support the researcher during analysis.

Introduction

Quantitative research tells us how often or how many behave in a certain way. However, when we want to broaden or deepen our understanding of how things came to be the way they are in our social world, qualitative research has no substitute[1]. In recent years, there has been an increase in number of qualitative studies been carried out by health professionals. Qualitative research usually produces large amounts of textual data in the form of transcripts and field notes. The systematic and rigorous preparation and analysis of qualitative data is usually time consuming and labour intensive[2]. To lessen this burden, it is important that health researchers become aware of the possibilities of using Computer Assisted Qualitative Data Analysis Software (CAQDAS) such as ATLAS.ti, MAXqda, NVivo and N6.

The dawn of CAQDAS was marked by the development of a program called Non-numerical Unstructured Data Indexing Searching and Theorizing (NUD*IST), which was specifically designed for qualitative data management in 1980's. NVivo, one of the current popular qualitative data management programs, has its roots in the NUD*IST. The developers of the software have described it as an improved and expanded version of the NUD*IST[3]. In his paper, Tom Richards, the co-developer of the program, justified the birth of NVivo in the face of the reputation and long history of NUD*IST. He stated that above and beyond NUD*IST, NVivo has features such as character-based coding, rich text capabilities and multimedia functions that are crucial for qualitative data management. Additionally, NVivo also has in-built facilities that allow people from different geographical spaces to work on the same data files at the same time through a network. Moreover, the strength of NVivo lies in its high compatibility to research designs. The software is not methodological-specific, it works well with wide range of qualitative research designs and data analysis methods such as discourse analysis, grounded theory, conversation analysis, ethnography, literature reviews, phenomenology, and mixed methods. For instance, Donmozoun et al. (Qualitative study)[4], Adengo et al. (Comparative study)[3], Gilmore et al. (Cross-sectional)[5] and Hill et al. (Systematic Review and Meta-Analysis)[6] used NVivo to manage data for their respective different study designs. Although CAQDAS has been available since 1984, very few qualitative health researchers (16%) report using it[7]. This may be due to the hassles that one has to go through to master the software and or the misconceptions that are associated with using CAQDAS[7]. While the issue of mastering CAQDAS has received ample attention, little has been done to address the misconceptions associated with the use of CAQDAS. In this paper, the author reflects on his own experience of using NVivo. The objective is to provide evidence-based implications of using qualitative data analysis software, so as to keep fellow health researchers well-informed, and consequently allow them to make informed decisions on whether to use CAQDAS or not. To make this practical, I will use one of the recent studies I have analysed using NVivo. Actual data management process using NVivo is presented first and is followed by deep reflection of the process.

Methodology

Data source

No specific fieldwork or data collection was commissioned for the current study. The reflections presented in this paper are by-products of the data analysis that the author conducted from 2011 to 2012 for a study titled 'Examining the differences in uptake of antenatal care between young and old women in Zomba, Malawi'. The aforementioned study was conducted to understand the differences in uptake of antenatal care (ANC) between young pregnant women (primagravida) and grown-up mothers (multigravida). It was generally a qualitative comparative study, involving poor villagers in Zomba district, Malawi. By the nature of the study topic, all the participants were pregnant women. In total, ten pregnant women were recruited from Zomba central hospital and interviewed right there at the hospital. The two data collectors went to the health facility for five consecutive days to conduct in-depth interviews. On each particular day, they recruited and interviewed the oldest and youngest pregnant women who attended the ANC (two per day). All the interviews were audio recorded, checked for quality and then kept in a password-locked computer that was only accessible to the researchers of the project. Ethical approval was granted by Malawi's National Health Science Research Committee.

Data management and analysis

Since the reflections presented in the current paper are based on the analysis of the aforesaid study, the whole process of analysis has been described in detail below. The intention of the study was to develop explanations direct from the data, hence grounded theory approach to analysis was adopted[8]. I didn't have much time to transcribe all the interviews, so I decided to use the analysis software that could also allow me to code both transcripts and audio files. NVivo software (version 9) was chosen because it met these requirements. Analysis started with the creation of a new project in NVivo, which I named "Care Seeking Project," and two source folders within the said project - for transcripts and audio

files. . I had ten sources in total for my project: five audio files in wmv format and five transcripts in Microsoft office word format. Although NVivo 9 manual that comes together with the software states that these formats are acceptable in NVivo; however, the program refused all the audio files with the message 'invalid format' appearing. After several attempts proved futile, I decided to downloaded the audio converter software, installed it in my laptop and converted the format of the audio files to mp3 which was eventually accepted by the NVivo. The transcript in word format were transferred without any problem. After all necessary files were transferred, the next step was coding, the process of putting together extracts (across documents) that are related to each other into basins called nodes[9]. Since analysis was driven by grounded theory principles, the first two transcripts and audio files were read and listened to in detail respectively and interesting excerpts were coded to free nodes. In particular, transcripts were thoroughly read and nodes were created in the process to house relevant excerpts or text from the transcripts. In NVivo, imported audio-file takes the form of an audio wave, which can be listened to and divided up into audio excerpts. The author followed NVivo audio coding process described by Wainwright and Russell[10]. Audio files were listened and relevant audio excerpts were coded to new and or existing nodes. From these nodes, five tree nodes were created (ANC, Children, Delivery, worries and pregnancy) and coding to these nodes continued with the rest of the documents. Two new nodes (God and Traditional cultures) were added in the process and one (worries) was combined with 'pregnancy' node. Content of the nodes were constantly reviewed by simply double clicking them. Coding stripes were also turned on to help manage the coding process by providing some insights for example, where the densest parts or coding are and so on. One of the important objectives of the study was to examine if demographic characteristics influence pregnant women's care seeking behaviour. To examine that, node classifications containing defined attributes for all respondents were created. When this was done, nodes associated with each source were created with the relevant details. Apart from coding to nodes, I was also able to connect ideas emerging from two or more sources using 'see also links'. Memos were also created for documenting my thoughts, doubts and insights that were emerging as I was going through the data. Annotations were created for both audio files and transcripts but they were particularly important when coding audio files because they were also acting as reminder or clarifications of audio excerpts. When the analysis process reached an advanced stage, a broader picture and or a visual representation of the data and progress of the work became a necessary prerequisite for development and testing of the theory. At this level, reports, queries, charts and models were created.

Findings: reflecting on the role of NVivo in qualitative data analysis

To begin with, just like manual analysis, I believe that NVivo can work well with most research designs and analytical approaches. There are no requirements relating to how the research should be designed[7]. I used grounded theory for my study but any approach could also fit. This basically implies that NVivo has little or no influence on the design of the research. In particular, however, I think that the presence of nodes in NVivo makes it more compatible with grounded theory and thematic analysis approaches. Moreover, the

nodes provide 'a simple to work with structure' for creating codes and discovering themes. Furthermore, NVivo also has a silent role which I managed to grasp during the analysis process of this study. The software has the potential to make the researcher more creative. Manual qualitative data analysis is so demanding and when you have, for instance, more than 1000 transcripts, chances that those sources can be read in detail are very low because of the copy-cut-paste burden posed by the traditional manual analysis system. Even though I only had ten files to work with in this project, NVivo still relieved me of the burden associated with manual coding. In a bigger project, this could provide room for investigators to focus on finding underlying themes, interpretation and theory instead of wasting time with copy-cut-paste manual coding processes. In addition, NVivo also ensures easy, effective and efficient coding which makes retrieval easier[11]. For instance, with NVivo I was able to link a paragraph from one source to another paragraph in either the same or another source and retrieve it with less effort. This task could have been very tough if I was using manual coding. Also, papers are difficult to manage and in manual coding the loss of even a single paper could cause serious damage to the project. In NVivo, all the sources are kept together under one roof. Although files are located in different places within the same project, the links that are created makes retrieval simple while in manual coding a researcher can spend long period of time searching for the missing papers or files rendering the process ineffective and inefficient. Likewise, when using NVivo it is easier to reshape and reorganise coding and nodes structure quickly[7]. For instance, deleting, copying, moving and combining nodes could be done without affecting the sources by simply clicking a few buttons. As already elucidated, during analysis, I first created five nodes then later combined the two and also add two. It only took me less than two minutes to do that. This could not be the case under manual analysis in which doing the same means an overhaul of the whole process.

Apart from this, NVivo also helps to improve accuracy of qualitative studies[11]. For instance, when I wanted to find out how many women with no child completed all ANC visits, manual searching took me over four hours while in NVivo it was only a matter of seconds and the results were also accurate and reliable. To ease the process of coding especially when you have many transcripts, same queries could also be used to identify common words and codes which could be the starting point for coding. This feature was not useful to me because my transcripts were few, so it was appropriate to read all of them thoroughly. Additionally, NVivo program could also generally facilitate accurate and transparent data analysis process[7]. If somebody may want to understand or follow what I did during the analysis process of the study above, it could be almost impossible if I had used manual analysis. The materials for manual analysis are best understood by the researcher him/herself and are usually difficult to sort. However, anyone who knows how to use NVivo can easily browse through my project and understand how the analysis of the data was done. On the other hand, although it has been put forward that NVivo is capable of accepting every research design; I still wonder whether we are not making the software fitting our research designs. When thinking about a project, we draw all the stages in our minds, hence there is the possibility of setting NVivo as our analysis tool at the beginning and then choose a design that would fit the software and then claiming that NVivo accept all designs. My

study was not exceptional to this drawback. I cannot 100% claim that NVivo fitted well into my study because I already had the software in mind when I was designing the project. It might be possible that I forced the package to fit my study or my study to fit the software. Either way is possible and dangerous as it might affect research results. Also, the idea behind using NVivo is to make qualitative data analysis easier[11]. However, this intention is to a larger extent defeated if we consider the tough time that one has to go through in order to learn how to use the software. The process is not only difficult but also time consuming. Although the software comes with a detailed tutorial on how to use it, the serious assumption of the tutorial is that the researcher has the basic understanding of the computer, which may not always be the case. I consider myself to have advanced computer skills but I struggled to learn the software, then what more with a researcher with limited computer skills?

Moreover, although NVivo 9 accepts most file formats, there are quite a number of them which it does not. From the recorder, my audio files showed that they were all in WMA format which NVivo claims to accept. As mentioned earlier however, several attempts to add the files into NVivo resulted in a message 'invalid format'. I had to download an audio file convertor to change the format of the files to mp3, which NVivo then accepted without problems. This was not an easy task at all and could present a major challenge to the users of the software who are not skilful enough. Furthermore, although it has been argued elsewhere that NVivo does not take over the analysis process from the researcher[7], there is still possibility that the NVivo structure could determine rules for specific procedures. For instance, the software already has nodes which is a leading sign to the researcher to split his/her data into those containers (nodes). Likewise, although features like queries have been applauded for their contribution during coding[12], it can also be argued that these features might serve to distance the researcher from the context of the data[7]. Coding based on query's searches can obviously dilute thickness of the data. Use of manifold synonyms might lead to partial retrieval of information. For instance, in my study I wanted to find out if distance is an important factor to ANC attendance, so I searched for the word 'distance'. I got a few results which I coded to a node. But when I went through the transcripts in detail, I realised that other words like 'the hospital is very far', which also implies distance, were missed by the query search. So the question is, should we entrust this responsibility to NVivo?

Conclusion

NVivo (and of course all CAQDAS) now forms important part of qualitative data analysis. Among others, NVivo saves researchers from 'time consuming' transcription and boost the accuracy and speed of the analysis process. This is not to say that qualitative data analysis programs are 100% perfect; just like statistical packages, CAQDAS too have drawbacks. The key message to take home is that unlike statistical software, the main function of CAQDAS is not necessarily to analyse data, but rather to aid the analysis process, which the researcher must always remain in control of. In other words, researchers must equally know that no software can analyse qualitative data. NVivo and all other CAQDAS are basically data management packages, which are there to support the researcher during the data analysis process.

References

1. Hancock B. An Introduction to Qualitative Research. London; 2007.

2. Pope C, Ziebland S, Mays N. Analysing qualitative data. 2000;320(January):5-7.

3. Adongo P, Tabong P, Azongo T, et al. A comparative qualitative study of misconceptions associated with contraceptive use in southern and northern Ghana. Front. Public Heal. 2014;2(September). doi:10.1080/13645570210146267.

4. Donmozoun TS, Sombie I, Meda N. What prevent women for a sustainable use of maternal care in two medical districts of Burkina Faso? A qualitative study. PanAfrican Med. J. 2014;8688:1-6. doi:10.11604/pamj.2014.18.43.2210.

5. Gilmore B, Vallières F, McAuliffe E, Tumwesigye N, Muyambi G. The last one heard: the importance of an early stage participatory evaluation for programme implementation. Implement. Sci. 2014;9(1):137. doi:10.1186/s13012-014-0137-5.

6. Hill J, Mello-guyett LD, Hoyt J, Eijk AM Van, Kuile FO, Webster J. Women's Access and Provider Practices for the Case Management of Malaria during Pregnancy : A Systematic Review and Meta-Analysis. PLoS Med. 2014;11(8). doi:10.1371/journal.pmed.1001688.

7. Mclafferty E, Farley A. Analysing qualitative research data using computer software. Nurs. Times 2006;102(24):34-36.

8. Strass A, Corbin J. The Basics of Qualitative Research: Grounded Theory Procedures and Techniques. California: Sage; 1990.

9. Richards L. Handling Qualitative Data: A Practical Guide. London: Sage; 2005.

10. Wainwright M, Russell A. Using NVivo audio–coding: Practical, sensorial and epistemological considerations. Soc. Res. Updat. 2010;(60). Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Using+NVivo+Audio-Coding:+Practical,+Sensorial+and+Epistemological+Considerations#0. Accessed December 9, 2014.

11. Bezeley P. Qualitative Data Analysis with NVivo. London: Sage; 2007.

12. Walsh M. Teaching Qualitative Analysis Using QSR NVivo. Qual. Rep. 2003;8(2):251-256.