

NVivo 8 and consistency in data analysis: reflecting on the use of a qualitative data analysis program

Cite this article as: Bergin M (2011) NVivo 8 and consistency in data analysis: reflecting on the use of a qualitative data analysis program. *Nurse Researcher*. 18, 3, 6-12.

Received March 19 2010; accepted November 15 2010

Correspondence to
Michael Bergin
mbergin@wit.ie

Michael Bergin is a lecturer in nursing at the Department of Nursing, School of Health Sciences, Waterford Institute of Technology, Ireland

Abstract

Aim Qualitative data analysis is a complex process and demands clear thinking on the part of the analyst. However, a number of deficiencies may obstruct the research analyst during the process, leading to inconsistencies occurring. This paper is a reflection on the use of a qualitative data analysis program, NVivo 8, and its usefulness in identifying consistency and inconsistency during the coding process.

Background The author was conducting a large-scale study of providers and users of mental health services in Ireland. He used NVivo 8 to store, code and analyse the data and this paper reflects some of his observations during the study.

Discussion The demands placed on the analyst in trying to balance the mechanics of working

through a qualitative data analysis program, while simultaneously remaining conscious of the value of all sources are highlighted.

Conclusion NVivo 8 as a qualitative data analysis program is a challenging but valuable means for advancing the robustness of qualitative research.

Implications for practice Pitfalls can be avoided during analysis by running queries as the analyst progresses from tree node to tree node rather than leaving it to a stage whereby data analysis is well advanced.

Keywords

Qualitative data analysis, software, computer program, research

THERE IS much debate for and against the use of qualitative data analysis (QDA) software in qualitative research.

Advantages include a single location for storage that provides easy access to material and the ability to handle large amounts of data with consistent coding schemes. Other advantages include improvements in the consistency of approach, assistance with team research and the ability to help in theory building (Weitzman 2000).

Disadvantages include the amount of time and effort taken to become proficient in using the program; prescriptive approaches to analysis with a reluctance to change categories of information once they have been developed (Robson 2002); and the tendency for the analyst

to take short cuts (Weitzman 2000). The issue of whether or not using the program distances researchers from their data (Bong 2002, Roberts and Wilson 2002) and thereby impedes or distorts analysis is widely debated (Bourdon 2002). Analogies of creating a 'Frankenstein's monster' have been put forward but have tended to be over-emphasised (Kelle 2004).

However, the qualitative researcher who does not use software beyond a word processor will be hindered in comparison with those who do (Miles and Huberman 1994), although some argue that the computer needs to be used as a tool purely for data administration and archiving and not for analysing the data (Kelle 2004). Therefore, in trying to make a decision about whether to use a QDA program, it is

worth bearing in mind what Gibbs (2004) suggests: it is not the computer that interprets the text but the person.

Aim

This paper is a reflection on the use of a QDA program and how it helps the researcher as analyst to identify consistency or inconsistency in data analysis. One of the central requirements in qualitative analysis is clear thinking on the part of the analyst. However, a number of deficiencies of the 'human as analyst' can obstruct analysis (Robson 2002). These include: data overload; first impressions; information availability; positive instances; internal consistency; uneven reliability; missing information; revision of hypotheses; fictional base; confidence in judgement; co-occurrence; and inconsistency (Robson 2002).

Robson explains that:

- Positive instances refer to a tendency to ignore references that are conflicting with beliefs already held and to place a greater importance on information that confirms them.
- First impressions refer to the influence that early input can have so that subsequent revision is resisted.
- Internal consistency relates to a tendency to discount new or unusual information.
- Missing information refers to when information is incomplete it can be devalued.
- Fictional base is the tendency to compare with a base or average when no base data are available.
- Co-occurrence is when data that co-occur are misrepresented as strong evidence for correlation.
- Inconsistency is when repeated evaluations of the same data tend to differ.
- Revision of hypotheses is a tendency to either over or under react to new information (Robson 2002).
- The other four deficiencies – data overload, information availability, confidence in judgement, uneven reliability – are dealt with in more detail in the reflection.

I will explore some of these deficiencies to try to explain inconsistencies during the coding process. Before doing this, I will present an explanation of the workings of NVivo, followed by a brief outline of the research study and how data were coded to free and tree nodes.

NVivo NVivo is a QDA software package produced by QSR International. It supports the analysis of qualitative data in five principal ways (Bazeley 2007):

- Managing and organising data.
- Managing ideas.
- Querying data.
- Graphically modelling the ideas and concepts that are being built from the data.
- Reporting from the data.

For my study, the time and effort spent in learning NVivo was of great value, since the study was large scale with two sample groups: service users and service providers. I did not spend time learning the program at the beginning of the study but rather as I progressed through its various stages: for example, when I was importing interviews and literature, creating case nodes, transcribing, coding, querying the data, creating models, and writing memos and propositions. However, I received one day of introductory training in NVivo and this gave me an appreciation of what the program had to offer.

NVivo is designed for qualitative researchers who need deep levels of analysis for small or large volumes of data. NVivo has always set the standard in qualitative analysis and NVivo 8 is no exception. For the first time, the software works with audio and visual material, and new tools mean deeper and more flexible analysis is possible (QSR International 2010). This feature was one of the reasons that influenced my preference for NVivo as it allowed me to import all digitally recorded interviews and then transcribe them. This made it easier for me to listen and re-listen to the interview recordings and check for accuracy. My interview transcripts were not word-for-word accounts, however, but summaries of what participants said. This approach is supported by a realist method – see 'Method' (page 8).

I also chose NVivo for pragmatic reasons: its licence was available at my work, ongoing training and support were available from an NVivo expert, and I could access the support remotely at any time using the internet. The facility to have an expert access my project remotely to view any problems or issues I had was of great benefit as I progressed my study. While I found the software to be user-friendly, the ongoing support and training allowed me to maximise my use of the program, thereby benefiting my study. It was this support that alerted me to the issue of inconsistency in my coding on. Additionally, NVivo allowed me to import my literature and link the literature and audio recordings to the findings.

Focus of study The study focused on the views of 26 service users and 28 service providers on current mental health service provision in Ireland. Its aim encompassed the following objectives:

- To analyse the views of service users and service providers on current mental health

service provision with reference to the theoretical and empirical literature on critical realism, gender and mental health.

- To identify and analyse data relevant to the development of a gender sensitive culture in mental health.
- To explore and analyse organisational issues that affect service provision as this relates to gender.
- To make recommendations with regard to mental health policy and gender sensitive service structures in mental health.

Method

I used a qualitative methodology based on Layder's (1993, 1998, 2006) adaptive and social domains theories. This methodological approach is underpinned by a realist meta-theory (Bhaskar 1989, 1998, 2008) and concentrates on 'interweaving' the macro (organisational, institutional) and micro (individual) features of social life against a backdrop extending across time and space of multiple social domains and their relationships to power (Layder 1998, 2006). This understanding acknowledges the richness, complexity and depth of the social world and avoids the reductive tendencies of symbolic interactionism and grounded theory, phenomenology and postmodernism (Layder 2006). Bergin *et al* (2008, 2010) convincingly argued for critical realism as a philosophical framework for this study and its relevance to issues of gender and mental health.

A realist approach is at variance to some of the differing constructivist traditions (for example, symbolic interactionism, ethnomethodology, phenomenology). These traditions all describe reality as being socially constructed, what exists depending on a person's interaction, interpretation and understanding of that reality (or multiple realities) and its socially produced knowledge (Guba and Lincoln 1994, Crotty 1998). Therefore, these approaches prefer word-for-word transcriptions to capture that entire social reality.

However, constructivist approaches have been criticised for failing to recognise that there is more to social reality than what is articulated through the language of individuals. It is, as Layder (1998) suggested, impossible to understand social interactions without a concern for system factors and how they intertwine with each other. Therefore, from a realist perspective, an understanding beyond what is articulated by individuals is required so that the social world is more fully understood, explained and evaluated.

Thus, in transcribing interviews it was sufficient for me to capture what participants said as

summaries and not word-for-word accounts. Also, I considered my literature base to be data. This is supported by the idea that prior theoretical ideas, concepts, models or propositions can be helpful in relation to theoretical sampling and theory generation (Layder 1998).

Coding the data Coding data in NVivo 8 involves the creation of nodes. A node is a collection of references about a specific theme, place, person or other area of interest (Bazeley 2007). NVivo 8 allows different types of nodes to be created, including free nodes, tree nodes, case nodes, relationship nodes and matrices. You create nodes during a coding process by reading through references or sources, such as interviews or focus groups, and then categorising this information. A 'free node' is a stand-alone node that has no clear logical connection with other nodes - it does not easily fit into a hierarchical structure. Free nodes can be useful when you begin coding and have not yet developed a node structure.

Tree nodes are organised in a hierarchical structure, moving from a general category at the top - the parent node - to more specific categories or child nodes. You can use these to organise nodes for easy access in a way similar to a library catalogue. You can convert a free node into a 'tree node' by moving it into a tree node folder.

Cases are nodes with attributes, such as participant type, age, occupation and duration in clinical practice. These can represent people, institutes, sites or other entities involved in your research. Relationships are nodes that define the connection between two project items, such as a doctor and nurse, or between two tree nodes, such as symptoms and diagnoses.

Lastly, matrices are the result of a matrix-coding query (see 'Querying your data') and the matrix cells created are nodes that can be used to explore and further code on the data.

Having imported my interviews into NVivo 8, I began creating free nodes. I read through all the transcripts and coded the references to broad themes or free nodes, creating 51 free nodes in total. A feature of qualitative analysis is that the creation of nodes (themes) becomes less frequent as you progress because data (references) begin to fit free nodes that already exist. The number of references coded to each free node varied from 1,126 coded to the free node 'care provision' to 15 coded to 'childcare'. There were some free nodes that had a small number of references and these were combined with other free nodes; for example, I combined 'childcare' with the free node 'parenting'.

Of the original 51 free nodes I created, I coded on 41 to create a hierarchical structure of tree nodes (parent and child nodes) organised into three key areas for gender awareness: sensitivity, ideology and knowledge. I found the software worked very well for coding the data and did not restrict my ability to create categories and sub-categories. I was quickly able to identify free and tree nodes that overlapped or had few references. The hierarchical structure made it easier and more convenient for me to view and get a sense of my findings and this compared more favourably to previous studies that I had undertaken where I used pen, paper and highlighter pens. Once I had completed the coding, I wrote a summary and propositional statement for each child node. This process spanned a year and demanded great concentration. When I had completed it, I had to check my statements and summaries to determine if my interpretations of the data were sufficiently supported by cases (service users and providers) and references. This led me to query my data.

Querying your data Queries enable you to probe your data, find patterns and pursue ideas (Bazeley 2007). You can save queries, re-run them through new data and track the progression of results. You can create and save the following queries in NVivo8:

- **Text search query:** this lists all sources that contain specified text. This can provide a quick way of coding your sources – for example, you can search for words and code the occurrences at a particular node.
- **Word frequency query:** this lists words and the number of times they occur in selected items. Seeing which words appear most frequently can help you to identify themes and concepts.
- **Coding query:** this gathers content based on how it was coded. For example, ‘show all the content where service users talk about their symptoms’.
- **Matrix coding query:** this creates a matrix of nodes based on search criteria (QSR International 2010). I used this query most in the study. For example, I queried the number of cases and references coded by participant type and sex. In doing this, I was able to view the number of cases and references coded at free nodes by the selected attributes and compare how I further coded to parent and child nodes. It is the results of this that inform my reflection below.

There are two further queries that can be created:

- **A compound query:** combines text and coding queries. It searches for specified text in or near coded content.

- **A coding comparison query:** compares the codings of two researchers or two groups of researchers. This is useful for teams that are interested in coding consistency and the questions posed by researcher disagreement.

Matrix-coding query I mostly used matrix-coding queries to determine the number of cases and references coded by the attributes ‘participant type’ (service user and provider) and ‘sex’. In so doing, I could view the number of cases and references coded at the free nodes by selected attributes and then compare how I coded on to the parent and child nodes.

Table 1 shows the node ‘responsive’. The first row shows initial coding to the free node and highlights that I coded 25 service users and a similar number of providers. However, coding on from here revealed significant differences between the numbers of cases for service users and for service providers. There were noticeable differences at many of the sub-categories or themes (child nodes) that were created during the coding-on process – see, for example, ‘diagnoses-based sensitivity’ and ‘insightful to parenting issues’. This discrepancy alerted me that I needed to review all service-user references at 30 of the 41 free nodes. This was a necessary and time-consuming process. However, the outcomes were positive as many of the sub-categories (child nodes) and their statements are now more supported by the data.

It is suggested that some qualitative researchers consider the process of counting references or cases to be at variance or misleading (Seale and Silverman 1997, Miles and Huberman 1994). I did this only to highlight how well supported my statements and summaries were for the two sample groups and to demonstrate how I had coded on. This was a valuable exercise for me as it displayed my inconsistency in coding on.

The majority of child node statements and their summaries were largely unchanged following this review (Table 2). This may be because I had read through and carefully considered service-user references at each node during all the stages of coding, and that I was aware of their content when it came to writing my statements and summaries. However, there was a small number of summaries that I wrote initially from the perspective only of service providers because of the limited number of references; this changed to both perspectives following the review.

Inconsistency in coding on: a reflection As mentioned above, when I discovered there were

Table 1 Matrix query before review of service user references

Responsive	Service user	Service provider	Male	Female
	25	25	17	33
Barriers to being responsive and sensitive to needs	6	15	9	12
Changes for service provision	8	1	4	5
Cultural awareness	0	7	3	4
Developing greater sensitivity, responsiveness and awareness	9	16	6	19
Diagnoses-based sensitivity	3	21	6	18
Equality for all	10	11	8	13
Immediacy and sensitivity of care	0	1	0	1
Insightful to parenting issues	3	19	4	18
Policy and gender sensitivity	8	15	5	18
Receptive to services	1	3	2	2
Roles and improving awareness and sensitivity	0	2	0	2
Service users' views of responsiveness of services	22	1	9	14
Socialisation of men and women	1	11	4	8
Structural and organisational issues	6	20	8	18
Sympathetic to men and women	2	16	5	13
Symptoms and services' understanding and sensitivity	4	17	6	15
Tolerant, receptive and sensitive to the needs of minority groups	0	17	3	14
Understanding the concept 'gender' and sensitivity of service provision	2	10	3	9

significant differences in how I coded on service-user references compared with that of service providers, I reviewed all service-user references at 30 of the 41 free nodes. There are a number of reasons that explain this inconsistency in coding from free node to tree and child nodes. I will address this using some of the deficiencies that qualitative researchers can encounter when they are analysts. These include 'data overload', 'information availability', 'uneven reliability' and 'confidence in judgement'.

'Data overload' I coded more than 15,000 references to the 41 free nodes. While I coded many to a number of free nodes, the volume remained and each reference at a free node (or multiple nodes) required consideration in how to further code on. This was demanding, as each node had to be worked through and required great concentration. This was particularly significant in relation to how I sequenced my interview

transcripts. Each node had service-provider references sequenced alphabetically followed by service-user references. This was challenging, especially as 17 nodes had more than 400 references, 10 had 700 or more, and two nodes had more than 1,000. Thus, at each node I had to work initially through the service provider references before progressing to service-user references. Sometimes, I concentrated less than necessary and I was selective at times.

This problem was further compounded by service-user interviews being transcribed verbatim in Microsoft Word. I could only view these transcripts and this compared poorly with service-provider interviews that I digitally recorded, imported into NVivo 8 and then transcribed in a manner that made them easier to read.

'Information availability' One of the advantages of the new tools available in NVivo 8 is that they

Table 2 Matrix query following review of service-user references

Responsive	Service user	Service provider	Male	Female
	25	25	17	33
Barriers to being responsive and sensitive to needs	20	16	13	23
Changes for service provision	23	1	10	14
Cultural awareness	14	8	10	12
Developing greater sensitivity, responsiveness and awareness	20	17	11	26
Diagnoses-based sensitivity	17	22	14	25
Equality for all	20	11	13	18
Immediacy and sensitivity of care	14	2	5	11
Insightful to parenting issues	7	20	7	20
Policy and gender sensitivity	16	16	9	23
Receptive to services	21	4	11	14
Roles and improving awareness and sensitivity	18	3	6	15
Service users' views of responsiveness of services	24	1	10	15
Socialisation of men and women	9	11	6	14
Structural and organisational issues	19	21	14	26
Sympathetic to men and women	20	17	13	24
Symptoms and services' understanding and sensitivity	16	18	12	22
Tolerant, receptive and sensitive to the needs of minority groups	4	17	4	17
Understanding the concept 'gender' and sensitivity of service provision	19	11	9	21

enable you to import, play, analyse and transcribe your audio and video files. This makes it easier to document and summarise what is said and allows you to re-listen to the audio. The digital recording can be viewed as your first script and typed transcripts can be a summary of what is being said. When you import an audio or video file, NVivo 8 creates a 'source' containing the media and an 'empty' transcript. This source creates rows that contain a time span and a corresponding typed content area. These rows can be created to reflect specific issues that arise as the audio is listened to. This allows large amounts of transcript to be easily viewed and I found this facility favourable for working with the data.

However, service-user interviews were not digitally recorded and were typed in Microsoft Word. The transcripts were then imported into NVivo 8. This resulted in a large amount of typed transcriptions that were difficult to view, and time

consuming and laborious to read. Thus, analysing and coding on were more challenging here than with service-provider interviews.

'Uneven reliability' The fact that some sources are more reliable than others tends to be ignored by the analyst (Robson 2002). On reflection, with no conscious intent to ignore, I focused more on service providers and what they were articulating to the detriment of service users. While my statements and summaries for each node remained largely unchanged following a review of service-user references, the statements and summaries are now more supported by the data.

While I was aware of the content of service-user references, there was a degree of passivity further compounded by data overload and information availability that contributed to selective encoding and a rush to believe that service-provider sources were more reliable than service-user sources.

‘Confidence in judgement’ One can place excessive confidence in one’s judgement once it is made (Robson 2002). This may have led me to fail to question my initial judgements. I based judgements for each node largely on service-provider references, partially because of the data sequencing in NVivo. Also, ‘uneven reliability’ tended to bias my initial judgements.

Conclusion

Qualitative data analysis is a complex process and demands clear thinking on the part of the analyst. However, researchers as analysts encounter many deficiencies during coding and in making sense of their data. Selective attention to some issues more than others, especially when overloaded with data, can detract analysts from widely and evenly distributing their attention and interpretation of the data. This creates a tendency for concentration levels to lapse and the possibility of the analyst becoming purely focused on the mechanistic aspects of analysis, especially when using a QDA program. Selective encoding and a rush to a way of thinking can also lead to some sources being viewed as more

reliable and others being ignored. My use of NVivo highlighted these deficiencies during coding.

The use of a QDA program such as NVivo 8 allows for consistent coding schemes and provides the analyst with the tools to query and audit the coding processes. In so doing, more robust interpretations of data can be achieved. However, analysts need to be aware of the demands placed on them in trying to balance the mechanics of working through a QDA program while simultaneously remaining conscious of the value of all sources in their studies. This was particularly evident from my coding on of the data. This pitfall can be avoided by running queries as the analyst progresses from tree node to tree node rather than leaving it to a stage where data analysis is well advanced. Additionally, the digital facility in NVivo 8 has the advantage of allowing the analyst to engage data aurally and visually. This compares more favourably with traditional approaches where data can only be visualised.

In conclusion, NVivo 8 as a QDA program is challenging but a valuable means for advancing the robustness of qualitative research.

Acknowledgements

I would like to thank Ben Meehan, NVivo Training, Consultancy and Support, Dublin; John Wells, head of the Department of Nursing, Waterford Institute of Technology, Ireland; and Sara Owen, dean of faculty at the University of Lincoln, UK, for their assistance

This article has been subject to a double-blind review and checked using antiplagiarism software

Online archive

For related information, visit our online archive of more than 6,000 articles and search using the keywords.

References

- Bazeley P** (2007) *Qualitative Data Analysis with NVivo*. Sage Publications, London
- Bergin M, Wells JS, Owen S** (2008) Critical realism: a philosophical framework for the study of gender and mental health. *Nursing Philosophy*. 9, 3, 169-179.
- Bergin M, Wells JS, Owen S** (2010) Relating realist metatheory to issues of gender and mental health. *Journal of Psychiatric and Mental Health Nursing*. 17, 5, 442-451.
- Bhaskar R** (1989) *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. Verso, London.
- Bhaskar R** (1998) Philosophy and scientific realism. In Archer M, Bhaskar R, Collier A et al (Eds) *Essential Readings: Critical Realism*. Routledge, London.
- Bhaskar R** (2008) *A Realist Theory of Science*. Second edition. Routledge, London.
- Bong SA** (2002) *Debunking Myths in Qualitative Data Analysis*. www.qualitative-research.net/index.php/fqs/article/view/849 (Last accessed: February 9 2011.)
- Bourdon S** (2002) *The Integration of Qualitative Data Analysis Software in Research Strategies: Resistances and Possibilities*. www.qualitative-research.net/index.php/fqs/article/view/850 (Last accessed: February 9 2011.)
- Crotty M** (1998) *The Foundations of Social Research: Meaning and Perspectives in the Research Process*. First edition. Sage Publications, London.
- Gibbs GR** (2004) Searching for text. In Seale C (Ed) *Social Research Methods: A Reader*. Routledge, London, 307-311.
- Guba EG, Lincoln YS** (1994) Competing paradigms in qualitative research. In: Denzin N, Lincoln YS (Eds) *Handbook of Qualitative Research*. First edition. Sage Publications, Thousand Oaks CA, 105-117.
- Kelle U** (2004) Theory building in qualitative research and computer programs for the management of textual data. In Seale C (Ed) *Social Research Methods: A Reader*. Routledge, London, 312-317.
- Layder D** (1993) *New Strategies in Social Research: An Introduction and Guide*. Polity Press, Cambridge.
- Layder D** (1998) *Sociological Practice: Linking Theory and Social Research*. Sage Publications, London.
- Layder D** (2006) *Understanding Social Theory*. Second edition. Sage Publications, London.
- Miles MB, Huberman AM** (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. Second edition. Sage Publications, London.
- QSR International** (2010) *NVivo 8 Tutorials*. www.qsrinternational.com/support_tutorials.aspx?productid=18 (Last accessed: January 9 2011.)
- Roberts KA, Wilson RW** (2002) *ICT and the Research Process: Issues Around the Compatibility of Technology With Qualitative Data Analysis*. www.qualitative-research.net/index.php/fqs/article/view/862 (Last accessed: January 9 2011.)
- Robson C** (2002) *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Second edition. Blackwell Publishers, Oxford.
- Seale C, Silverman D** (1997) Ensuring rigor in qualitative research. *European Journal of Public Health*. 7, 4, 379-384.
- Weitzman EA** (2000) Software and qualitative research. In: Denzin NK, Lincoln TS (Eds) *Handbook of Qualitative Research*. Second edition. Sage Publications, Thousand Oaks CA, 803-820.

Copyright of Nurse Researcher is the property of RCN Publishing Company and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.