Using Software to Open up the ‘Black Box’ of Qualitative Data Analysis in Evaluations

The Experience of a Multi-Site Team Using NUD*IST Version 6

CHIH HOONG SIN
Disability Rights Commission, UK

The lack of transparency in the process of qualitative data analysis has led to the suspicion that findings may not be robust enough to be used as evidence. Much qualitative data analysis software (QDAS) contains functions that facilitate the demonstration of reliability and validity, although the process is not straightforward. Given the scale and complexity of many evaluations, QDAS is a powerful tool that can be used to manage and analyse vast amounts of qualitative data. Its use in an evaluative context, however, can be fundamentally different from using it for other forms of qualitative analysis. This article explores the use of NUD*IST version 6 (N6), a widely used QDAS, to manage data from the qualitative aspects of a large-scale multi-component evaluation involving a large, multi-site team. It demonstrates that research and logistical imperatives can have concrete impact on the way technology is harnessed to produce qualitative findings, hence inevitably influencing resultant evidence.

KEYWORDS: evidence; qualitative data analysis software; reliability; validity

Introduction

The way in which qualitative approaches are operationalized and qualitative data analysed in terms of evaluative research have been largely undocumented (for exceptions see di Gregorio, 2001; Sin, 2003a, 2003b). The procedures of qualitative data management and analysis in relation to evaluation can often be perceived as shrouded in mystery. This is unhelpful as the lack of transparency can often lead to the charge that qualitative data are not robust enough to be used as ‘evidence’, particularly if such ‘evidence’ is to be used to inform policy-making. A document
published by the UK Cabinet Office in August 2003 on the assessment of quality in qualitative evaluation noted that:

... there are widespread concerns about quality. There is also shared interest in issues such as ‘rigour’; the need for principles of practice to be made manifest; the importance of sound or ‘robust’ qualitative research evidence; and in the relevance and utility of research. (Spencer et al., 2003: 4)

Quite often, the use of ‘grounded theory’ is simply mentioned as a shorthand for a particular approach to data analysis, as if the mere mention of grounded theory explains the complexity of analysis. In fact, Lee and Fielding (1995) found that while it is common for researchers to mention a ‘grounded theory’ approach, most do not actually adhere to the procedures laid out by Glaser and Strauss (1967). Thus, the process through which qualitative data are turned into evaluative evidence has become a ‘black box’ that needs unpacking.

It has been argued elsewhere that qualitative data analysis software (QDAS) has the potential to open up the ‘black box’ of qualitative data management and analysis (see Richards, 2005; Sin, 2004) as well as demonstrating reliability and validity (Richards, 2004). However, simply because a particular QDAS has such functions does not necessarily translate into them being used appropriately or effectively. It is quite common to find researchers citing the use of particular QDAS as if the mere mention of a particular software is a ‘quality guarantee’ (MacMillan and Koenig, 2004). However, the extent to which QDAS is able to help demonstrate reliability and validity of evaluation findings is dependent on the way it is put to use. Hence the attention should be focused on the ways in which QDAS has been employed.

QDAS and Evaluation

The way QDAS is deployed for evaluation research differs from its use on most other qualitative research in two significant ways. First, most of the existing literature on qualitative research is written from the perspective of lone researchers. Evaluation research, on the other hand, is characterized by teams of researchers. The scale and complexity of much evaluative research mean that diverse expertise is called for from a range of individuals possessing different skills.

There is an urgent need for more work exploring how QDAS is used in team contexts, particularly given the observation by Gibbs et al. (2002) that increasingly sophisticated technology has enabled the undertaking of larger qualitative studies and the fact that the same piece of technology can be employed differently to yield different results in different contexts (see Mangabeira, 1996).

Second, much existing literature on qualitative research using QDAS report on stand-alone qualitative projects. Evaluation, in contrast, tends to involve mixed methods, particularly those that are large-scale and complex. The use of QDAS for qualitative data management and analysis is arguably different in a mixed-method context since the qualitative data gathered for evaluations are not usually
analysed in and for themselves, but rather in complement to the quantitative data
gathered for other components of the evaluation.

These factors influence the way in which QDAS is used for evaluative research,
and, inevitably having a direct and indirect influence on the data created,
interpreted and reported; and ultimately the ‘evidence’ produced.

This article reflects on the experience of using one of the most common
QDAS packages: NUD*IST version 6 (N6) in a big team, multi-site large-scale
evaluation project, demonstrating that its use for evaluation often reflects a series
of pragmatic decisions. N6 is based on the coding or tagging of sections of text-
based data (e.g. interview transcripts and notes) that assists the researcher in
analysing and exploring qualitative data (see www.qsrinternational.com for more
information). N6 allows the users to identify themes that are of interest. These
themes may be pre-defined or may be emergent during the process of analysis.
The resultant codes can be organized into a structured thematic ‘coding tree’
(although they do not have to be).

The Study

The project reported here is a large-scale multi-component national evaluation
of a community and neighbourhood renewal initiative in Britain commissioned
by a government department. This article reports on the setting up and use of N6
during the first research phase of this three-year evaluation.

Choosing N6

Prior expectations for N6 Although members of the evaluation team did not,
at that time, have much experience in the use of QDAS, the use of N6 had been
explicitly stated in the proposal submitted for the evaluation. It was almost as if
the mention of the use of N6 added a sheen of scientific rigour that is required to
justify the use of a variety of qualitative techniques which are ‘soft’ (see Gibbs et
al., 2002).

Another reason for the use of N6 emerged after the evaluation had been
commissioned. Unrealistic expectations of what the software was capable of led
to an implicit expectation that N6 would somehow analyse the data at the touch
of a button, metaphorically speaking. However, as Weitzman and Miles (1995),
among others, have noted, computers do not analyse data; people do!

Related closely to this was an additional expectation that N6 would be able to
test a series of hypotheses more or less unambiguously. As Kelle (1997) pointed
out, this amounted to the use of N6 within a hypothetico-deductive research
strategy to test, empirically, statements about the exact relationship between
two defined variables. The temptation for doing so in evaluative research is
particularly strong due to the imperative of evaluation, in general, to establish
causality. However, Kelle cautioned that using ‘fuzzy codes’ on qualitative data
to test precise hypotheses runs the risk of confusing incompatible research logics.
This, as Carvajal (2002) pointed out, is one of the common misconceptions of
what N6 (or any other QDAS package) is for.
Not only was N6 supposed to help analyse the considerable amount of data, it was also expected to do so quickly. There was a belief that using N6 would save time. While there is some truth in this assertion, for instance in speeding up the mechanical aspects of data management, whether or not N6 really does save time depends on a host of other factors such as the nature of the research, the characteristics of the data and the types of analyses required (Carvajal, 2002).

Further assessment of the suitability of N6  Concern over the fact that the use of N6 has been undebated led eventually to more attention being paid to the assessment of its suitability for this particular evaluation. There were two issues for consideration. First, should any QDAS package be used at all? It does not always require dedicated software to manage and help analyse qualitative material. For instance, WORD and ACCESS have been used to perform certain aspects of qualitative data handling (see Meyer et al., 2002; Nideröst, 2002).

If QDAS was necessary, a second question of ‘which package would be most suitable for the purposes of this evaluation?’ should be answered. The suitability of individual packages needs to be assessed not merely in the context of the qualitative component of the evaluation but also by taking a holistic view of the entire evaluation strategy and looking at how the use of a particular package aligns with the range of research instruments, as well as with the quantitative components of the evaluation.

As a result of such assessment, N6 was retained for use. N6 sits well with the other modes of managing data. For instance, an ACCESS database had already been set up to log information pertaining to case-study area characteristics. Table-based processing in spreadsheets and statistics packages is particularly well-suited for integration with N6. The programme contains sophisticated ‘import’ and ‘export’ facilities. The former enables the user to ‘import’ tables of data from packages such as EXCEL, SPSS and ACCESS, and ‘convert them into base-data or demographic coding of documents or case nodes (rows in the table) by values of demographic variables (columns in the table) that become sub-trees of base data nodes’ in N6 (Richards, 2002: 207).

Additionally, outputs from N6 can be exported for further analysis in programmes such as SPSS that can perform more sophisticated quantitative analysis. Numerical data generated by the results of matrix node searches could also be exported. These include information such as the numbers of documents or numbers of passages (‘text units’) coded under each node (see Bazeley, 2002a, for a detailed worked example of how such functions have been used in real research situations).

These functions in N6 bring about more than just technical integration and ease of data management in complex projects such as evaluations; they enhance the ability of evaluators to achieve higher levels of analytic integration of qualitative and quantitative data sources (Bazeley, 2002b), thereby ensuring the findings of evaluation are more robust. As Bazeley (2002a: 241) noted, programmes such as N6 allowed ‘the integration of textual and statistical analyses in a project using mixed form data, while retaining the full hermeneutic richness of that data’. Nonetheless, there is recognition that the capability of N6 to link up with quantitative software is contentious (e.g. Hesse-Biber, 1995) and can be abused.
In addition, the project management aspect of N6, including the ability to produce an ‘audit trial’ (Barry, 1998: paragraph 9.1), is highly attractive in the context of this evaluation, given its size and complexity. For example, the memo function within N6 can be used to chart a detailed account of the development of an analysis plan. It can be used for documenting information on ways of interpreting and interrogating data; decisions made about coding; explanations of how decisions were made or modified.

Furthermore, given that data collection was to be conducted in four research phases over three years, the ability of N6 to manage data collected at different time periods is hugely important. The ease of retrieval and management that N6 lends to the process of analysis means that, as a long-term evaluation unfolds and new issues emerge (e.g. due to policy changes), it is easy for the evaluators to revisit original documents, interview transcripts or observation notes and perform new analyses should it be necessary.

N6 also facilitates work to be conducted by multiple researchers based at different sites, even those using different hardware (PC or MAC). Researchers can work on coding separately and then combine the work done at stages using the ‘Merge’ function (see Tagg, 2002, for lengthier discussion of various coding and merging strategies). This was absolutely vital in the case of this evaluation since it involved a team of 15 people during the first research phase employed by two organizations, with several individuals working on this project from home.

### Getting Ready to Use N6

#### Wider Training Needs: Knowledge
There were two main challenges facing the evaluation team at this stage. First, everyone had to be brought up to speed within a relatively short period of time. Second, apart from imparting the theoretical, conceptual and methodological know-how, what was perhaps equally (if not more) important was ensuring that each team member had the confidence to be able to use N6 effectively and to be clear about his/her roles and responsibilities within a team context.

Therefore, instead of launching directly into training sessions on the use of N6, it was felt necessary to conduct a series of training sessions on (1) the background and aims of the policy initiative, situating this within the wider sociopolitical context; (2) the theory and practice of evaluation in general, and this piece of evaluation in particular; (3) the background to undertaking qualitative research, subsequently focusing on the particular tools to be employed in this study; (4) an overview of the entire evaluation project, outlining how each part is intimately connected to the others.

#### Wider Training Needs: Teamwork
In addition to the knowledge aspect of training sessions, a separate series of team-building sessions were conducted. These were felt to be crucial if a big team was to function properly. At these team-building sessions, individuals were encouraged to discuss frankly their fears and hopes of being part of the evaluation team;
what their strengths and weaknesses were; and the types of support they would require.

**Setting up a Coding Tree in N6**

**Situating Coding within the Wider Evaluation**

A first version of a coding tree was constructed based on the understanding of the wider research context and the need to harmonize the data and analyses for the range of research tools employed for the evaluation. Research tools have been designed in such a way that there is a harmonization of concepts (or rather of themes). A series of topics is covered using different research instruments. Each research tool was designed to allow the evaluation team to bring all necessary data to bear on a point from a variety of sources. The triangulation of data enhances the robustness of findings and also ensures that the battery of research instruments form a coherent research strategy.

The need to situate the use of N6 firmly within this holistic structure of the evaluation was one of the major factors influencing the design of the coding tree. Certain significant themes were pre-identified and helped inform the guidelines for conducting semi-structured interviews to a greater or lesser extent. Thus the analysis of the qualitative data yielded is not purely inductive from the start, but has been guided by more or less well-defined questions. This approach represents a significant departure from the conventional ‘grounded theory’ approach used in qualitative research (Glaser and Strauss, 1967).

**Situating Coding within the Context of Teamwork**

A second main influence on the coding tree came from the consideration of using N6 within a big-team, multi-site context and the implications for managing and coordinating its use. There were early attempts at constructing a coding tree using a series of high-level abstract codes such as ‘process’, ‘structure’, ‘organizational behaviour’. However, it soon became evident that having a coding tree made up of such high-level abstract concepts was not workable in the context as the interpretation of such concepts was very loose.

Inter-coder reliability is an issue to be contended with when there is more than one person doing coding (Bourdon, 2001; Ford et al., 2000). While it is impossible to ensure perfect inter-coder reliability, there is nonetheless a need to ensure that the variance in interpretation is minimized as far as possible. With this in mind, a different type of coding tree was called for.

The resultant coding structure utilized a series of themes that were predefined to a certain extent by the structure of the evaluation and by the other research tools utilized in the study. Contrary to the debates surrounding the extent to which grounded theory and its link to QDAS have influenced the way qualitative research has been taking place (e.g. Coffey et al., 1996), the codes used in this study were neither purely emergent nor entirely grounded. Instead, they were descriptive ‘themes’ that were drawn from the evaluation’s main lines of inquiry. While the evaluation team approached the use of N6 with theoretical
preconceptions, revisions and modifications were possible and expected once they had immersed themselves in the data.

There were several immediate benefits to using such a coding structure. First, the themes reflected the guidelines for conducting the semi-structured interviews. Therefore, given the semi-directive nature of the initial interviews, it was only logical that such themes would be important at the coding and analysis phase. Second, these themes had been informed partly by an extensive literature review in the area and partly in discussion with the entire evaluation team and an expert panel. Thus the understandings and salience of these themes had evolved through a process of discussion and negotiation in the months prior to the actual coding exercise.

Shared understandings of the themes had developed by involving each and every member of the team, thus minimizing inter-coder variance at the stage of using N6 to code. As Crow et al. (1992) argued, differences in interpretation had to be negotiated.

Training to Use N6

A ‘Hands-on’ Approach and Discussion

When the coding tree was set up, a training session was held which began with a reminder to the team of the relation between N6 and qualitative research, before moving on to the mechanical aspects of the coding functions. As different members of the team had different levels of computing proficiency, the hands-on session was extremely useful.

Working with actual data also made the use of N6 real and enabled additional forms of checks on inter-coder reliability. For instance, every individual was told to use the coding tree to code an identical segment of text. The team then got together to discuss how each member of the team had coded particular text units. This encouraged discussions around why particular text units were interpreted in certain ways. In several instances, it helped refine the coding tree, for instance in making existing codes even less ambiguous or to add in new codes that had not been foreseen.

This roundtable discussion was also invaluable in another way: it brought out the importance of contextual, non-textual, information pertaining to the local organization and conduct of interviews that inevitably influenced the way each member of the team interpreted particular text segments. Thus, seemingly unambiguous text units could sometimes be interpreted differently.

The discussions allowed such contextual, subjective experiences to emerge and allowed the team to come to a consensus on how to deal with them in terms of coding. As Ford et al. (2000) argued, ‘increased inter-rater reliability could be produced by the interpersonal dynamics and is certainly to be expected as a common vocabulary develops’.

The importance of non-textual information also meant that the team agreed that it would be better for each team member to code the transcripts of interviews that they themselves had conducted using an agreed coding structure. As
Tagg (2002: 282) noted, however, managing a team effort in this way meant that ‘each researcher cannot use the power of the software to develop nodes and a node structure as they code; the coding and its structure must be predefined’. To prevent the enterprise from being too directive and to allow for the potential of emergent codes, the coding tree was also designed in such a way that each team member had a preassigned area for developing ‘personal coding trees’.

**Practising in a Simulated ‘Real-Life’ Scenario**

As a result of the training session, a revised guidance document was prepared. A new dummy project file was sent to each team member, complete with a revised coding tree and the same set of sample transcripts. Team members had one week to practise with this dummy project file.

This exercise served three purposes. First, it enabled individuals to grow more confident in the use of N6. Second, the use of N6 only becomes real when one works with real data (Carvajal, 2002). By practising in a simulated ‘real-life’ situation, individuals come to a better understanding of how they interpret and code text units. They also get a better feel for how they work best with N6. Third, this trial run also allowed for the possibility of further refinement to the coding tree and to allow each individual to assess the adequacy of the coding tree to their sets of documents. This spurred them to consider how they might use their assigned personal coding tree nodes, and what occasions would call for them to be used.

At the end of the trial week, a final version of the coding tree was produced within a final master project file that was sent to every team member. Together with this, a revised and updated guidance document was also sent to everyone, this time including the feedback from the trial week. With all these in place, team members were told to work on their transcripts for real and to send in their project files for merging on a regular basis. This broadly exemplifies the first of four strategies proposed by Tagg (2002: 280–2) for managing team projects using the merge function.

**Conclusion**

Various authors have expressed the view that one of the possible reasons why ‘few texts in qualitative methods even discuss the origins or organization of codes, categories or labels by which data is classified, and even fewer discuss the problems of index management’ (Richards and Richards, 1992: 6, quoted in Cannon, 1998) may be due to the fear of attracting harsh criticism from the scholarly community about the often less than ideal ways in which research was actually operationalized.

However, users of QDAS should take the opportunity to make their research process more transparent as this serves to highlight the fact that the way data are manipulated, managed and represented is part and parcel of the social construction of evidence used in evaluation. This article, for instance, has demonstrated that the functions contained in a particular research software, combined with the way in which the use of that software is managed in a multi-
site team context, can have concrete impact on the procedures adopted during the data analysis process. These have impinged particularly on the formulation of codes with which to interrogate the data, thus irrevocably influencing the way evidence is constructed and understood. This has massive implications for the way we understand and consume the products of evaluation. By opening up the ‘black box’ of qualitative data analysis, evaluators can learn from the mistakes of others and improve on ways of managing and analysing data.

Acknowledgements

The author would like to acknowledge the support from Jack Cattell, Sarah Edwards, Nilesh Goswami, Neil Graver, Jacque Mallender, Alex May, Kerry McCarthy, Ben Monks, Andy Nichols, Chris O’Leary, Duncan Stacey, Michael Stanworth and Naser Turabi. The encouragement from Stuart Campbell, Lyn Richards and Clare Tagg is also gratefully acknowledged. The author takes sole responsibility for the views expressed in this article.

References


Sin: Using Software to Open The ‘Black Box’ of Qualitative Data Analysis


CHIH HOONG SIN is Head of Information and Research at the Disability Rights Commission, Great Britain. He is interested in issues relating to equalities and diversity, having been involved in research on ageing, race and ethnicity, as well as disability. He has designed and conducted a range of mixed-method evaluations of social policy interventions covering substance misuse, neighbourhood renewal, health, children and young people, and criminal justice for a number of government departments. He has an ongoing interest in issues relating to research methodology. Address for correspondence: Head of Information and Research, Disability Rights Commission, 3rd Floor, Fox Court, 14 Grays Inn Road, London WC1X 8HN, UK. [email: Chih.HoongSin@drc-gb.org or Chihhoong@hotmail.com]