Using qualitative data analysis software (QDAS) to assist data analyses

Pat Bazeley

Translational Research and Social Innovation group, WSU

Western Sydney University

Email: pat@researchsupport.com.au

This is a submitted version of this publication. Publication details are:

Bazeley, P. (2018). Using qualitative data analysis software (QDAS) to assist data analyses, in P. Liamputtong (Ed.), *Handbook of research methods in health social sciences.* Singapore: Springer Nature. doi:10.1007/978-981-10-2779-6_108-1

Abstract

Qualitative data analysis software (QDAS) has much to offer the health researcher. Software facilitates efficient management of qualitative and mixed methods data through a variety of tools to organize and keep track of multiple data sources and types, and of the ideas flowing from those data. Coding tools provide structure to the categories and themes evidenced in the data, allowing for rapid retrieval of information. Increased depth and rigor of analysis is facilitated through capacity to search and interrogate the data sources using a combination of coding and other data management tools. Questions can be asked about how often and how different categories or themes are expressed by different groups within a sample, or within different contexts or times. Similarly, experiential data might be compared for those with different measures on health-related variables. Relationships between different aspects of experience (or attitudes or feelings, etc.) can be explored and/or verified using coding queries, through a range of visual displays, or through statistical analyses using exported coding information. Such queries can be limited to one type of data, or multiple types of sources can be imported, coded, and analyzed together. Linking tools are used throughout to connect reflective thoughts to the data that prompted them, or interim results to the evidence that supports them. Explanations of these processes are illustrated by figures and examples.

Keywords

qualitative, software, analysis, mixed methods, coding, visualizing, theory building

1. Introduction

Various forms of software to support qualitative data analysis were developed for public use in the 1980s and 90s. These software developments followed closely on the rapid developments that occurred in use of qualitative approaches to research that began a decade earlier and were supported by the revolution occurring in computer science with the shift from mainframe to personal computers. While statistical software to analyze numeric data was widely adopted without question in the quantitatively dominated research climate that prevailed at the beginning of the digital era, adoption of qualitative software as a legitimate tool to assist analysis of textual data has been a much slower process—and continues to be resisted in some quarters. The majority of early programs focused on 'code and retrieve' as a primary tool to assist analysis, because this is what computers were good at doing. The capacity to query connections in coding was also there as an extension to code retrieval, but was often ignored. This was partly because querying required a higher level of technical skill in the user, but also for some, the use of logic based queries was (and is) seen as a poor substitute for the 'intuitive' connections arising from deep immersion in data, manifested in the work of skilled qualitative researchers. Qualitative software, in its current form, is a product of the developments that have taken place in both computer technology and methodology in the 21st Century; it provides a much wider and ever-expanding range of tools to emulate the tasks involved in qualitative analysis, yet its heritage in those early 1980s developments, with their emphasis on coding as a primary analytical task, is still very much evident.

One of the issues faced by developers of software for qualitative analysis was, and is, the diversity of qualitative methodologies, and the inherent complexity of the analytic and interpretive processes associated with those methodologies—processes that are not readily reduced to linear algorithms, such as occurred with statistical analyses. This complexity requires flexibility in the way software tools function, which in turn engenders complexity in software design—and for the user, a relatively steep 'learning curve' to achieve effective use of the software for her particular purposes (Gilbert, 2002). This further inhibits adoption by some researchers using qualitative methods. Nevertheless, adoption is now widespread, and use of software for analysis of text and visual data is becoming more or less expected in many fields.

Just as the qualitative 'revolution' challenged and changed foundations for research from the 1970s, so also a mixed methods revolution has been changing research practices, primarily since the 2000s (Mertens et al., 2016). There has been some adaptation in statistical

software, but it is qualitative software that is really taking up the challenge of developing ways to merge quantitative and qualitative data and analyses. Further challenges in a digital era of big data and social media are being met by both statistical and qualitative analysis programs. Software of both types is constantly evolving, as developers work to meet the rapidly evolving landscape of research methods in a digital age.

Although technology plays an increasingly important role in the health sciences, health is essentially about human populations and the individuals that make up those populations, with all of their physical, mental, emotional, and spiritual qualities (Huber et al., 2011; Liamputtong, in press). Used wisely and with sensitivity, qualitative data analysis software (QDAS) can assist health researchers in their work, insofar as it is focused on the human interface of health care. It does so by offering a flexible suite of tools to use in managing and working through qualitative (and mixed methods) data, aiding analysis and interpretation in the process. Users of QDAS find coding, memoing, linking, querying, and visualization tools that help them to work with, reflect on, and see their data in new ways. It can assist, but not substitute, for deep immersion in data and the skilled recognition and interpretation of linkages and patterns across data by the researcher that are the hallmarks of ground-breaking qualitative analysis. Additionally, use of software ensures more systematic and transparent analytic processes, with conclusions not only supported by evidence, but by evidence and a chain of thinking that can be traced. Ultimately, however, the analytic interpretation remains in the researcher's skilled hands.

Software is often dismissed as something that is useful only at the latter stages of a project, when analysis is in full swing. I propose in this chapter to take the reader through the main stages of conducting a research project, and to demonstrate that QDAS has a role to play from start to finish.

2. Establishing the goals of analysis

If we are to consider how software can assist health researchers achieve their goals for analysis, a first step must be to establish what those goals might be. And immediately one is faced with the diversity of health research, just as health care and health services are diverse—a reflection of the breadth of what health itself entails. Research goals can range from the conceptual to the practical; from measurement-oriented assessment of the benefit of a drug or surgical intervention through to exploring the phenomenological nature of particular health-related experiences (which might, coincidentally, be of drug or surgical intervention);

from a focus on policies, organizations and services to a focus on ordinary people and their daily lives.

The role (and value) of software for analysis has its starting point right here, at the point of establishing goals for analysis. What will this project be designed to achieve? What, and whose, purposes will it serve? QDAS contributes in three ways at this pre-analytical point: through recording reflections, mapping preliminary ideas, and reviewing and analyzing literature.

Keeping a record within the software of one's musings over purposes and goals is helpful insofar as the act of writing assists in building clarification. Having a written journal recording ideas and steps taken additionally provides the basis for an audit trail relating to project development, something that becomes useful when justifying choices made and explaining directions taken, during a later stage of reporting. Applying some codes to this document as it is being written assists during analysis and writing with locating particular ideas, and it generates starter codes to apply to literature or participant data. For those new to QDAS, this provides a gentle introduction to using software.

Modelling (mind-mapping) ideas about the topic being studied also helps in clarifying those ideas, and with project planning, by pointing to those things for which data will be needed (**see also Mind Maps in Qualitative Research)**. Items entered in a mind map or conceptual model become represented as codes that will capture data, links drawn between items in an early conceptual model become assumptions to test (Figure 1). Ideas sparked through mapping processes are recorded in the project journal, again assisting clarification and for later reference.

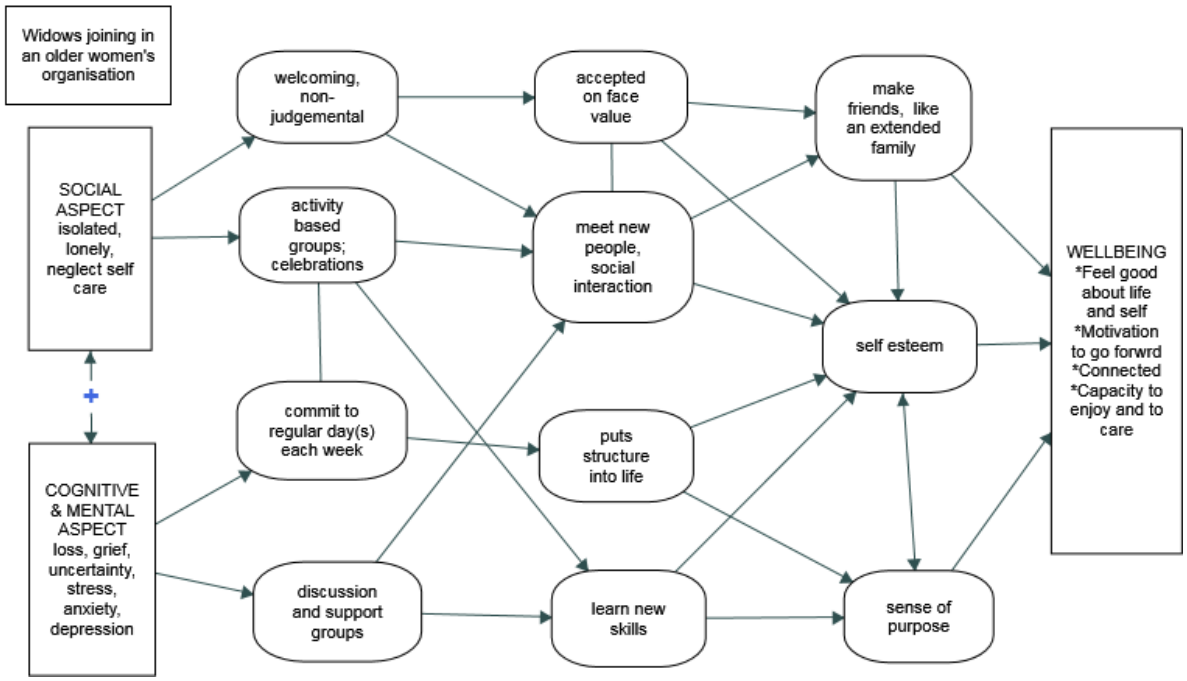


Figure 1 Conceptual model: how older women's groups benefit the social and mental wellbeing of widows

Most researchers now use bibliographic software to manage the literature relating to their research. That literature can be selectively imported from the bibliographic tool into the qualitative software, with abstracts, notes, metadata, and where available, the original .pdf sources. Once imported, those materials can be coded interactively, or searched using selected terms to identify relevant passages for consideration (rather than a search simply identifying whole references, as occurs within the bibliographic software). Alternatively, the content of articles might be rapidly scanned, aided by predictive auto coding tools that are available to varying degrees in different qualitative programs. Working through the literature on a subject is, of course, a key tool for refining objectives for the research and for identifying concepts and frameworks that, in turn, become tools to guide both design and analysis. When literature is incorporated into a project file, passages within it can be linked to new data, contributing to building a connected web of knowledge. Alternatively, queries can be used to compare what is being revealed through this project with established understanding as expressed in the literature. For example, professional understanding of caregiving for an aged person as a set of instrumental tasks is inconsistent with families' perceptions of caregiving as something more akin to psychosocial care; similarly discharge planners' views of home care needs can be very different from patients' perceptions of home

care needs, with such inconsistencies leading to problems of noncompliance (Bowers, 1989). (Coding and querying processes will be explained below.)

3. Selecting data for analysis

Traditionally, qualitative health researchers have relied heavily on interviews or, to a lesser extent, focus groups to provide research data. Diaries, journals, and observations also provide data for some, while visual data (video, images, drawings), and documents (e.g., medical records, nursing notes) offer less frequently accessed forms of qualitative data. Researchers are also exploring the potential of network, geographic, and social media sources (**see chapters in the Innovative Research Methods in Health Social Sciences section**). Health services and evaluation researchers taking a more quantitative or mixed orientation have relied largely on surveys and questionnaires, sometimes including some open-ended questions within those (**see Evaluation Research and Surveys and Questionnaires**). Increasingly, health researchers are broadening their traditional methodological approaches to combine these multiple sources and types of data, in what has become known as multimethod or mixed methods research (**see chapters in the Multi, Mixed Methodology and Collaborative Research section**) .

Relatively recent developments in QDAS ensure that almost any type of data can find a place in a project managed using software. As well as text data, most programs allow for importing video and image data in a variety of formats, several allow for recording geocodes (e.g., manually, or from GPS) and for displaying these in map format, some for incorporating geographic images as a base on which to 'map' other visually displayed qualitative data, and some now facilitate importing information directly from websites and/or social media platforms (including associated metadata). Additionally, survey data that includes open-ended responses can be imported directly from either Excel or on-line databases such as Survey Monkey or Qualtrics. Basic demographic data for participants has always been considered useful in qualitative projects; in mixed methods projects this is likely to be extended to include categorical responses and scaled data. QDAS usually deals with this demographic information and also any other quantitative variable data relating to participants or cases in the research by treating it as variable or attribute data which is then associated with related non-numeric data for those participants or cases. Finally, for those using ethnographic or other methods where data are continually added to ongoing field notes, records can be created and maintained within the software.

4. Managing data for analysis

Data management is rarely discussed in texts, and yet it is a crucial element in analysis, especially for the complex kinds of data that might be part of a qualitative project or a mixed methods project. (Miles & Huberman, 1994, p. 43) observe that "qualitative studies, especially those done by the lone researcher or the novice graduate student, are notorious for their vulnerability to poor study management". Approaches to and requirements for data management should be considered *before* data are gathered for a study, because these can impact on what can be done with those data during analysis—how they can be sorted, compared, and queried (Bazeley, 2013). In particular, some forms of data (e.g., focus group data, survey data) require special preparation to access the features of software that facilitate linking demographic, categorical, and scaled data with qualitative case data (Bazeley & Jackson, 2013).

Being able to trace and identify the source of evidence used, as well as the context for that evidence is important for the transparency and transferability of results. Keeping track of sources, and of segments of data from those sources, during the process of coding and analysis is facilitated in QDAS through the use of folders, descriptions, and labelling (Figure 2). Different kinds of sources are placed into folders or document groups, so that it becomes easier to see what data are available within the project. Contextual and reference information about specific sources can be recorded in descriptions, linked memos (a good place for associated field notes), or in associated variable data. And, whenever a particular segment of data is retrieved, the software indicates the source of that segment and provides the option to link directly from the segment back into its original context to assist in its interpretation (Figure 2).

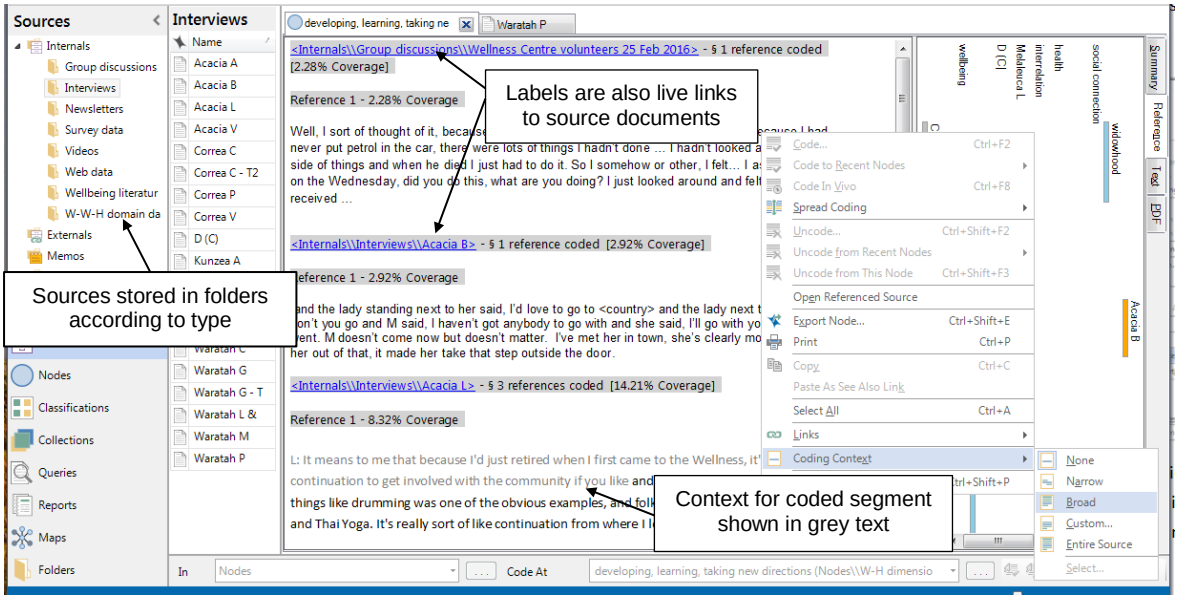


Figure 2 Keeping track of sources and context for coded segments

Managing sources in a way that allows for their analysis both as separate entities *and* as a common body of knowledge has always been a problem for qualitative researchers dealing with voluminous and messy data. QDAS assists this process by providing multiple systems for sorting and 'cutting' data, in ways that facilitate both targeted or comparative analyses.

Folders used for sorting and viewing sources become useful at the analysis stage as well. Some types of analyses are more or less appropriate with specific kinds of data, and so folders allow for scoping (focusing) a query to a specific group of sources.

Qualitative and mixed research projects are often structured around *cases*—the participants or other entities that exemplify the phenomenon being studied (Bazeley, 2013; Yin, 2014). NVivo facilitates the development of a dedicated structure for holding together diverse sources of data about each of the cases—the units of analysis—in a project, thus facilitating case-based analyses. Data for each case are *coded* to specifically designated categories, which means that data relating to a particular case can include one or multiple sources of the same or different types, and/or parts of sources. Thus, a participant's case data might include a combination of interviews along with their contribution to a focus group and their contributions to a meeting as discerned from the minutes of the meeting. In addition, attribute data (demographic, categorical, or scaled variables) relating to the case would be entered or imported and attached to that case's code, which means it is automatically applied to all data coded to that case, even if added later. Attribute values can then be used as a basis

for comparative analyses (as described in a later section). The alternative, for QDAS that do not offer a specialized case structure option, is to use the general coding structure to create a parent code for cases with a set of sub-codes. Each sub-code would then represent a case to which all relevant material is coded. In this situation, however, attribute (variable) data can be attached only to whole sources.

Sets are created in QDAS to hold together aliases for collections of sources, or of codes (or in NVivo, also a combination of sources and codes). Aliases update as the material they represent is updated. Sets are useful, for example, for separating sources created in different phases of a longitudinal data collection ($Time_1 \ldots Time_n$), or for designating different categories of contributors to a set of cases (e.g., when interviews are held with mothers, fathers, doctors, and the target child for each child [the case] with juvenile diabetes). Making sets of codes can be a useful strategy for identifying concepts that 'hang together', perhaps making for a more comprehensive or abstract category. The particular value of sets in analyses is that the software will treat the contents of a set as a single item, facilitating their use in comparative queries, or again, for scoping queries to a particular set of data.

Using these structural tools in combination will allow, for example, for a comparison of what different people [identified by sets] had to say about the anxiety [achieved by scoping the query to a particular code], over time [identified by sets] for children with juvenile diabetes [the cases] (Figure 3). Or, the anxiety [code used for scoping the query] of specific children [cases] or groups of children [identified using attribute values] might be compared over time (sets), based on the combined input of those who are connected to each case.

	Time 1	Time 2	Time 3
Child			
Mother			
Father			
Doctor			

Data in cells is scoped to include only text/data segments coded *anxiety*.

Figure 3 Using data management tools to support a complex comparative query

While folders, cases, and sets can be set up at any stage during a project, it is nevertheless important to have considered what they might look like at the start of data collection, especially where these impinge on the way data are prepared (what formatting is required), and how they are labelled. Good management facilitates good analysis; it also

prompts the qualitative researcher toward clarity and transparency, with potential impacts for ongoing design as well as analysis.

5. Reviewing and exploring data

Simply reviewing and/or exploring each data item as it is gathered is traditionally recommended as a first step in qualitative analysis (Bazeley, 2013; Liamputtong, 2017). In the early stages of a project, it can be found necessary to make adjustments to data collection procedures, based on these early reviews. Reviewing the whole of a source again, before detailed work is begun, helps to ensure specific elements within the source will be seen in the context of that whole. Many researchers prefer to read through and annotate a paper copy at this stage, although using software to add notes on or linked to specific points while reading through is not only possible, but has the advantage that those initial impressions are not lost to later analysis. Some also find it helpful to summarize the key points learned from a new data source; these can be recorded in a memo linked to that source. Another alternative is to use the mapping tools in software to create a visual diagram representing the narrative or perhaps the interconnected ideas being gleaned from the data source during this exploratory phase.

Of course, not all projects have data coming in one piece at a time, to allow the kind of work just described. Qualitative data can come in larger volumes, especially when sources other than interviews are being used. It is still valuable, however, to obtain an overview of what these new data are saying, before delving into detail. For this task, QDAS offers the options of either a global word frequency count, or targeted word searches. A word frequency count will identify the most commonly occurring words—the overview—and present these as a summary or in a visual display. Each line in the summary then provides a direct link from the listed word to the passages where it was found, and from there to the broader context of that find (e.g., the paragraph in which it appeared), or to the source documents, allowing for a deeper exploration of any words of interest, and the concepts they represent. These strategies are useful, also, for scanning the literature, as shown in Figure 4.

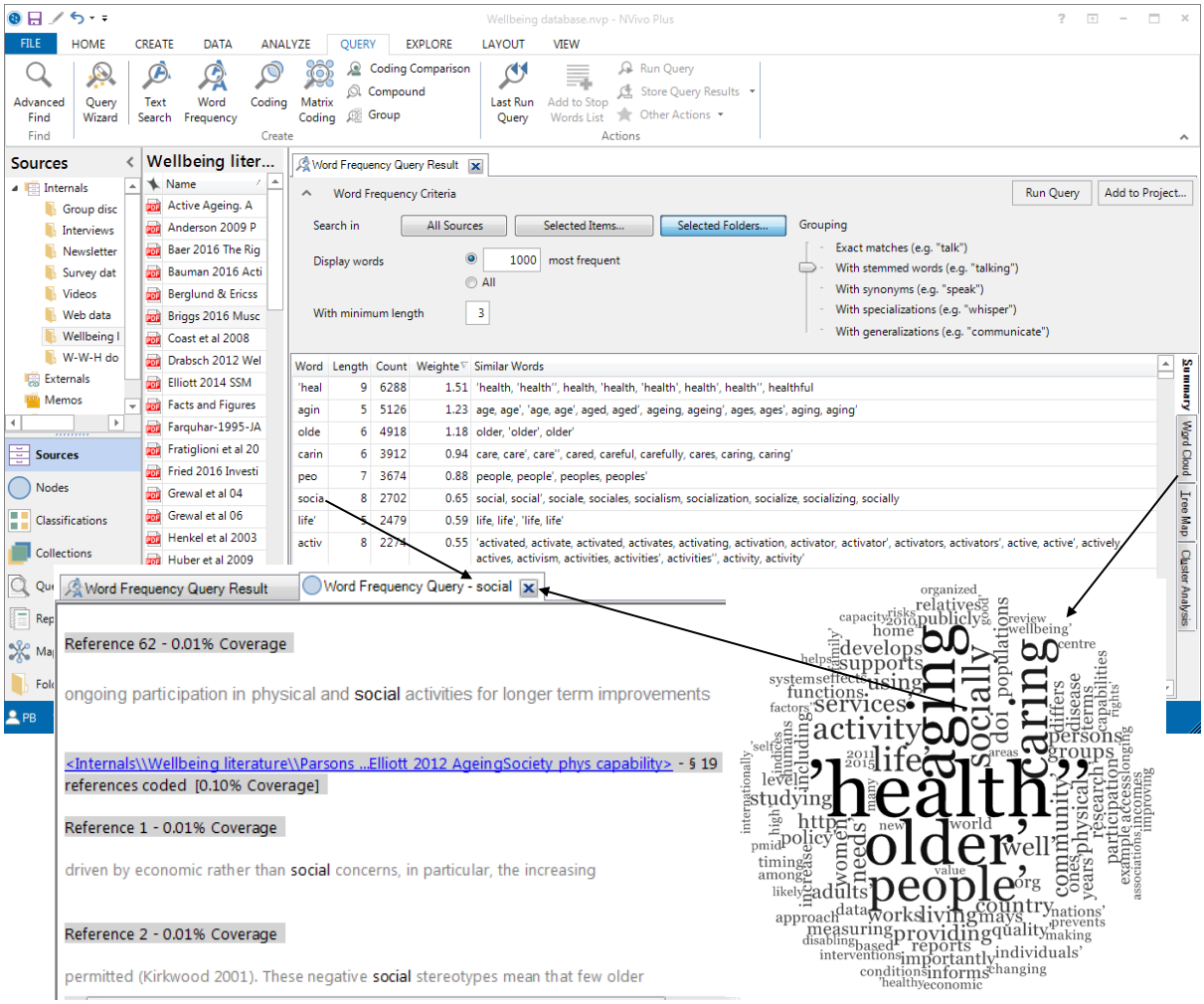


Figure 4 Word frequency query with links to associated text

Recording ideas as they occur and reflecting on what is being learned is useful at any stage of the analysis process, and this time of initial review and exploration is a good time to establish lasting habits for doing so. Ideas and reflections can be recorded either in a general journal for the project, or in memos linked to specific sources—the choice depends on what is being recorded about what (e.g., insights about the project as a whole, or its methodology, or reflections about the particular source). In the process of writing these reflections (of either kind), it is useful to link them to the source text or items that prompted them. In NVivo, a specific segment of data (or literature) that prompts an analytic or reflective thought is copied and pasted *as a link* within the journal or memo document at the point where the reflection is written. When the link is activated, it takes the researcher back to the copied evidence, showing it in its original context. In MAXQDA, separate memos for each passage are attached directly to the paragraph containing the passage that prompts the reflection, thus both programs achieve the same goal of connecting reflective thoughts to the evidence that

supports them. Similarly, connections can be recorded linking from text (or other detail) in one item to text (or other detail) in another (e.g., from a data item to relevant literature; from one reference to another; from part of a picture or video to a discussion about that component or event).

Journal or memo entries are optionally coded to categories used also for coding data. This has the benefit of allowing the researcher to easily retrieve not only data on a particular topic, but also reflections on that topic, potentially with links back to source data that provided the basis for the reflections. It overcomes any problems of trying to remember whether that note was made in the project journal or a particular memo.

Decisions made about the way the research is being conducted and about the direction of the analysis are also usefully recorded throughout the project (applying time stamps to these entries is a good idea). Some researchers choose to record these methodological steps in a separate journal, others use the general project journal. This builds an 'audit trail' for the project—a record that explains what was done and why it was done that way (Carcary, 2009; Richards, 2014; **see also The Nature of Qualitative Research**). This will help the researcher show the pathway taken in reaching the project's results and conclusions and allows for a more transparent write-up, for the benefit of the eventual readers.

6. Sorting and coding data: what are my data about?

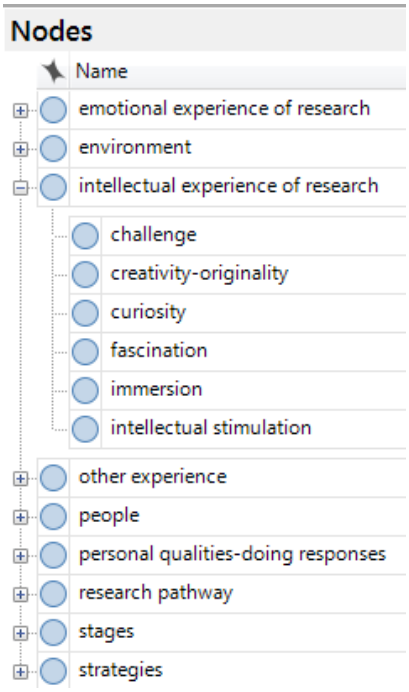
As much as making connections between data items the ideas in those data items is an important aspect of qualitative analysis, there is no question that coding is the basic tool that underpins most forms of qualitative analysis (Strauss, 1987). Coding is essentially a form of indexing that allows the researcher to find all the passages in their sources that relate to a particular topic, with the topic being represented by a label (the name for the code; **see Content Analysis, Thematic Analysis**). Coding, therefore, is an analytic task that connects data to ideas; it requires a level of interpretation by the person applying the codes (Bazeley, 2013). During the coding process, the researcher is also making connections between those ideas (these should be noted in the journal, to be further explored). Codes might be topic-based, contextual, or more abstract and conceptual. One of the simplest ways to move from descriptive coding to more abstract coding is to keep asking "Why am I interested in that?" (Richards, 2014). The codes developed in this way have significance beyond the immediate source, making them more analytically useful, although more descriptive codes retain their value also in recontextualizing more abstract concepts.

When coding was done manually, researchers would either write labels in the margin next to relevant passages, make copies of their documents and then cut them into sections that were then sorted into piles for each code, or create index cards that listed where relevant passages could be found, with a summary of what was in those passages. All of these methods imposed limitations on the researcher, either in terms of retrieving all that could be known about a particular topic, or more particularly, in sorting and connecting passages to explore or demonstrate a relationship between codes (e.g., a patient's mood when at work, compared with his mood at home; or, to identify the nature of a child's response to bullying at school, and then the impact of this on their sibling relationships). Because manual coding systems tend to limit the level of detail one can build into a coding system, they foster the use of more global codes in order to capture all of what is happening in a passage in one code, with a consequent tendency to rely on simple thematic analysis.

Coding with QDAS involves selecting passages and identifying what that passage is about. Each element of what is 'going on' in a passage is captured with separate codes, and each code represents a single concept. Thus, responses or actions that signify improvement in, say, a medical condition might be coded for the type of response and/or action, descriptively for the particular condition being referred to and for the circumstances where the improvement became evident, as well as for the more abstract category of 'improvement in condition', with each of these aspects being picked up by different codes, applied to the same passage of text. This has two consequences: firstly, all material relating to any particular code can be readily retrieved (with the source identified and accessible for each retrieved passage), allowing the researcher to explore in some depth what that concept is about—to see data in terms of the category rather than the original sources; and secondly, connections between codes, in terms of passages that are coded by particular combinations of codes, can be readily found using the query functions in the software. This means, for example, that the connection between a 'what' and a 'how' can always be re-established, such as how and where improvement was demonstrated, with relevant passages retrieved. More than that, however, connections between the same 'what' and different 'hows' can also be explored. Thus, using the examples given above, if codes for where a mood is being experienced are applied at the same time as codes for the type of mood being experienced, the software can sort out text to illustrate and compare which moods are being experienced where. Or, text about a child's response to a bullying experience can be associated with that child's way of interacting with siblings, and perhaps also with responses, say, for children living in different family circumstances being compared.

The way in which a QDAS coding system is structured has implications for how well a researcher can use the codes to explore and interrogate her data. The most effective coding systems, in terms of facilitating querying, are those that categorize codes by what kind of thing they are about, in a taxonomic system (rather like the way plants are categorized), that is, its branching structure is not theoretically determined. Thus, for example, different events will be listed as subcodes within one 'tree', emotions will be listed in another, as will times, places, or who is involved in particular events or emotions or whatever else is being coded (Figure 5 shows partially expanded examples for two projects). In an analytical sense, it means that any code can be associated with any other code, without limitations, with queries being able to find particular links as required, or patterns across groups of codes. In a purely practical sense, this means that a coding system is kept manageable (subcodes are not repeated across different trees; each can be a subcategory of only one 'kind of thing' and therefore has its own place). This type of structure helps the researcher to 'see' what her data are about—what kinds of concepts are involved. An added benefit for a supervisor of students engaged in qualitative research is that the structure and content of the coding system efficiently conveys the clarity and depth of a student's ongoing analysis.

Becoming and being a researcher



Dimensions of wellbeing for older women (one tree within a more extensive coding structure)

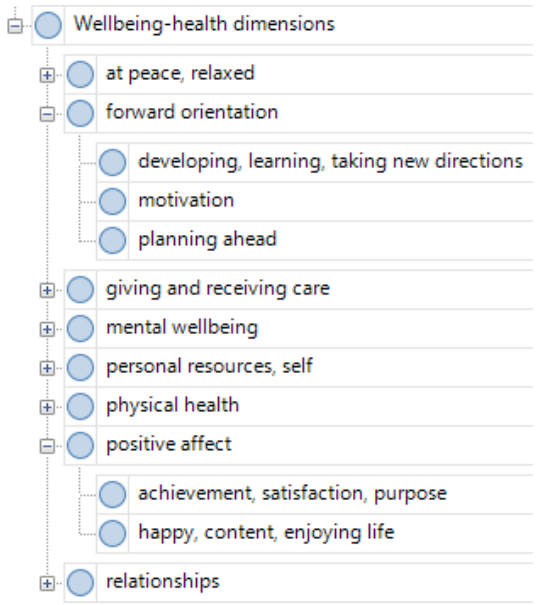


Figure 5 Sample coding structures for two projects

As codes are being moved in the process of structuring a coding system, increasingly attention is turned from the individual sources to retrieving, reviewing, and rethinking the concepts being worked with in a project. Descriptions will have been provided for codes when they were first created, or perhaps as further data are being added to them and further clarification is needed, or they are provided now as they are being moved around. During this structuring process, further memos about the coding process are likely to be added to existing journals or memos, or new memos are created and linked to specific codes, with each holding information or thoughts about that particular category, its nature, and what it might be associated with (to guide future queries). Consideration is given to what 'job' each code is doing, and how it might be used in later analyses. Codes that have only a few passages coded to them can be revisited and compared with others representing similar concepts, to see if they might be combined (if so, this warrants a note in the description, and/or in the methodological journal). Others might benefit from being split—coded on to different categories instead of, or in addition to, the original categories. As with text from sources, any thought-provoking segment found in the coded data can be linked to a note in a memo. When the linked segment is retrieved, it will be shown in its original (document or image) context. All of this activity (and interpretive thinking) is building an ever-deepening understanding of the data in the researcher's mind. As the coding, connecting, and memoing processes are proceeding, the researcher will sometimes use query tools offered by the software to check on aspects of coding. For example: has text coded for emotions also been coded for what gave rise to those emotions? This can be checked simply by retrieving the codes for various emotions and reviewing coding stripes. Was coding applied to all the times when patients expressed satisfaction with a service? A text search will locate all the times satisf* or related words (e.g., good OR lik* OR happ*) occurred in the text, and retrieve these showing the surrounding paragraph, so that coding can be checked. Even better, the search can be set to exclude passages already coded with satisfaction by using a query that combines a text search with a coding query.

Visual charting and mapping tools also continue to have a role as coding proceeds. Charts provide an overview of the coding for particular sources; coding stripes show the dominance and combinations of codes used in coding a source; coding stripes also show those codes that intersect with a code being reviewed; and simple conceptual maps created by the researcher serve as a visual aid in capturing ideas about connections in data, including literature. Visual tools based on coding are useful in an exploratory sense. For example, a coding comparison diagram shows the cases (or sources) coded at one or another or both of

two concepts, prompting thoughts about the relationship between these concepts. Thus, Figure 6 suggests the possibility of a strong relationship between social connection and engaging in physical activity in that most participants attending ‘Wellness Centers’ for older women spoke of both, and no one spoke of activity without also speaking of social connection. Alternatively, a cluster analysis of codes, based on commonality of words used, suggests relationships between pairs of codes that might warrant further investigation, as well as, in this case, pointing to the possible dimensionality of the concept of wellbeing for older women, based on the grouping of codes (indicators) used to describe it by those attending Wellness Centers (Figure 7).

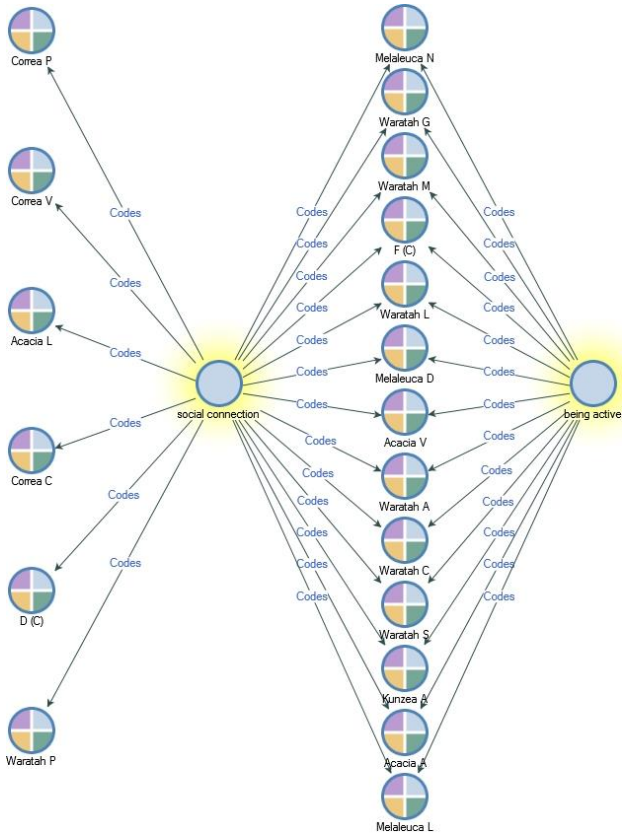


Figure 6 Coding comparison diagram showing cases coded at social connection and being active

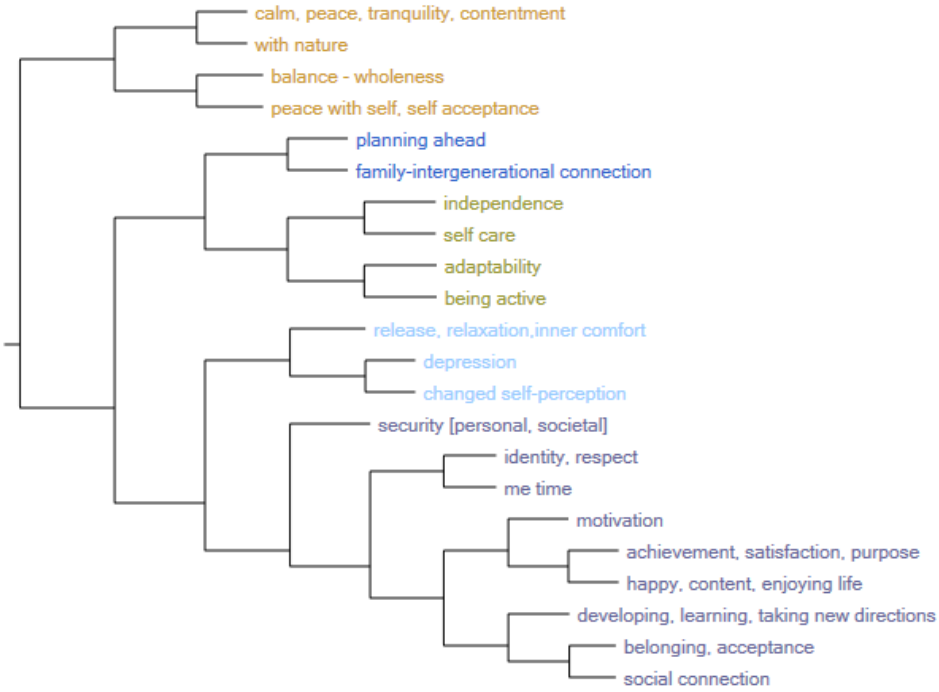


Figure 7 Clustering of codes used to describe wellbeing by older women

A final step, before interrogation of data starts in earnest, is to review the content coded at each coding category, firstly just to check that all the content coded there 'belongs' (checking consistency is a way of assessing coding reliability, without confounding the situation by having someone with a different perspective try to duplicate the coding process). More usefully, from the point of view of analysis, a review and summing up of each code clarifies both focus and boundaries for that concept, provides the descriptive beginnings of a report of results (helping to solve writer's block), and brings to light issues for further exploration—lines of investigation that will take the analysis beyond description of 'themes'. These summaries might be recorded in memos linked to the codes, with the issues for further investigation noted in the project journal. Alternatively, record the summaries directly into Word, using headings to identify each so that the navigation pane (document map) can be used to quickly locate them when further information becomes available.

7. Investigate and interrogate

The qualitative researcher is developing analytic ideas throughout coding and related processes, and particularly as a review of coded data is undertaken. A point is reached, however, where the focus shifts to a more deeply interpretive phase, where the researcher seeks to not just see, but to understand the data and what can be learned from them, to answer

the research questions and meet the purpose of the study. Actually, reviewing those research questions is advised at this stage because questions often shift to some degree through the course of a project. Once reviewed and realigned, the questions provide a guide to the direction of these further investigations, although some flexibility is advised as interesting leads can appear ‘late in the day’ (although perhaps these should be set aside for later investigation, especially when a deadline is involved).

Some incidental queries using the software might have been conducted along the way, as coding and structuring of the coding system was proceeding. At this stage, however, a process of deliberate and systematic querying is more likely, as patterns and connections in the data are investigated in the search for answers. Comparative analyses are good ways to start interrogating the data. Qualitative text (or other data) is compared for different subgroups of the sample, defined by values of demographic or other quantitative variables that have been stored as attributes attached to cases. In NVivo, a matrix coding query is used for this purpose (other QDAS offer equivalent tools). This provides the analyst with two kinds of data—the frequency with which different subgroups discuss the topics or experiences or issues being considered, and what is actually said about each of those by each subgroup. For example, Figure 8 compares what is said about some aspects of wellbeing for those who have different roles at Wellness Centers. Comparative analyses such as these are directly (descriptively) useful where comparative research questions were asked; analytically they are of value in prompting further investigation, such as why it is that this group had a different pattern of responses from that one—was it because…? Comparative analyses sometimes also reveal (sub)dimensions within concepts, when sampled groups have different ways of talking about the same concept. For example, communication patterns attributed to doctors might be compared according to whether they were described by nurses who do or who do not rate their approachability as an important issue, to reveal tacit differences in the way communication is perceived (e.g., two-way versus one-way) by these groups of nurses. These kinds of comparative queries are, therefore, useful for both exploratory analysis and for more directed analyses. Additionally, they can be extended to answer more complex three-way questions combining information from cases, sets, and codes, as described earlier.

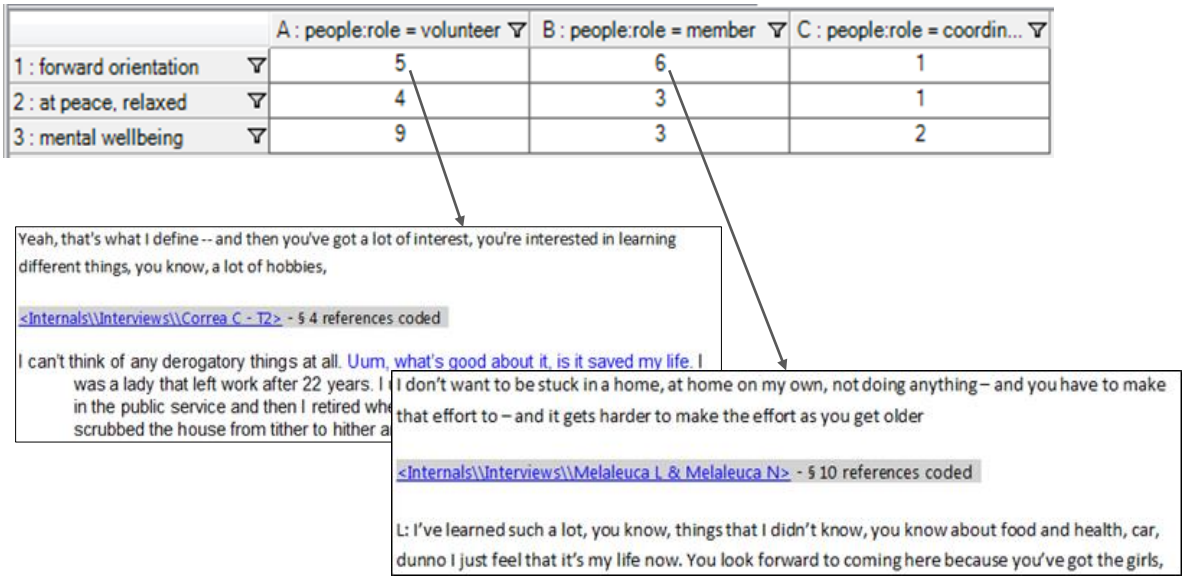


Figure 8 Comparison of some ways in which wellbeing is described, for people playing different roles in Wellness Centres

The matrix coding query is useful also for investigating patterns of relationships between groups of codes. For example, different (reported or observed) patient in-hospital experiences might be examined in relation to emotional responses of those patients, or perhaps in relation to patterns of adjustment or recovery once discharged. For a project evaluating strategies used to build adherence to guidelines for healthy eating (or exercise, or medication use), the way in which specific strategies were implemented can be reviewed in relation to whether they were considered helpful or not in terms of observable outcomes, as well as to how patients responded to each, and under what circumstances they were helpful—or indeed, some combination of these.

The way a coding system was structured is critical if it is to allow for asking a variety of questions, such as these—in the latter evaluation example, segments of texts describing responses to strategies would have been coded for the strategies being referred to, and in separate codes, whether they were observed to be useful, how the patient responded, and what the circumstances were at the time, with each of these codes being grouped in different 'trees'. This might appear, at the time of coding, to be unnecessarily fracturing the data, but the consequence is that it allows for flexibility in asking questions of the data when it comes time for analysis. Had strategies been listed in just two trees, as being helpful or not, then it would be difficult to ask further questions about other responses or influences.

Particular and/or more complex combinations of specific codes (or codes and attribute values) suggested either in earlier memos based on impressions when working with the data,

or through the kinds of visual and pattern analyses just described, might be assessed using regular coding queries, which will return all text satisfying the criteria set for the query. As for results of matrix queries, it is then up to the researcher to interpret patterns within the data that are revealed through the query.

Reporting from a qualitative or mixed methods study is best developed as the study proceeds, initially using the in-project document/memo system and the program's visual tools for mapping concepts and relationships, but then using a word processor alongside the qualitative software to record descriptions, insights, quotes, and helpful visual displays. The software thus provides an 'evidentiary database' (Yin, 2014) to be drawn on to support claims in the report.

8. Conclusion and Future Directions

What should be evident from the description given of these various processes for working with data when using software is the intensity of the way in which the researcher interacts with her data, putting to rest any possible claim or thought that using a computer might create distance between researcher and data (Jackson, in press). Qualitative researchers using software find it contributes to the rigor and transparency of their research processes and especially to the depth of analysis they are able to achieve. The software does not provide neat answers to the research questions, rather it is a tool that facilitates working with the data in ways that will provide the evidence needed for the researcher to make judgements and reach conclusions that are supported by data. Ultimately, of course, the responsibility for the depth and quality of the analysis and interpretation of results lies with the researcher.

In an increasingly digital world, qualitative and mixed methods researchers will also increasingly adopt digital technologies to aid their research work. And the imperative of business survival means that software developers will continue to develop new tools to assist that work. Qualitative software has moved progressively into managing an ever-expanding range of data types. There has been progressive development also in the kinds of analysis strategies that are supported, with the focus shifting most recently to automation of coding and now to mixed methods analysis, including basic statistical operations. Programs are handling ever-increasing volumes of data, which in itself requires new strategies for analysis. These trends in development will bring new opportunities, but also pose (or re-pose) threats to those who see such moves as eroding the 'real', 'intimate' character of qualitative research. Software *will* continue to support small-scale, intensive research; at the same time it will increasingly develop to capture new, volume-oriented markets. The future scenario will offer

choice, but that choice will not necessarily be ‘either-or’. Predictive coding strategies incorporated into software are improving in their capacity to apply machine learning, but perhaps of more interest are the moves afoot to develop and capitalize on ‘citizen science’ and crowd-sourcing strategies for hand-coding large volumes of data (Adams, 2016; Williams & Burnap, 2015) that are then fed into software used to assist with the (researcher-directed) analysis of those data.

References

Adams, N. A. (2016). A crowd content analysis assembly line: Scaling up hand-coding with text units of analysis. *Social Science Research Network (SSRN)*. Retrieved from http://ssrn.com/abstract=2808731

Bazeley, P. (2013). *Qualitative data analysis: practical strategies*. London: Sage.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2nd ed.). London: Sage.

Bowers, B. J. (1989). Grounded theory. In B. Sarter (Ed.), *Paths to knowledge: innovative research methods for nursing* (Vol. #15-2233, pp. 33-59). New York: National League for Nursing.

Carcary, M. (2009). The research audit trail - enhancing trustworthiness in qualitative inquiry. *Electronic Journal of Business Research Methods, 7*(1), 11-24.

Gilbert, L. S. (2002). Going the distance: 'closeness' in qualitative data analysis software. *International Journal of Social Research Methodology, 5*(3), 215-228.

Huber, M., Knottnerus, J. A., Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., . . . Smid, H. (2011). How should we define health? *British Medical Journal, 343*(d4163), 1-3. doi: 10.1136/bmj.d4163

Jackson, K. (in press). Where qualitative researchers and technologies meet: lessons from interactive digital art. *Qualitative Inquiry*. doi: 10.1177/1077800417731086

Liamputtong, P. (2017). Making sense of qualitative data. In P. Liamputtong (Ed.), *Research methods in health: Foundations for evidence-based practice* (3rd ed., pp. 421-436). Melbourne, AU: Oxford University Press.

Liamputtong, P. (in press). Health, illness and wellbeing: An introduction to social determinants of health. In P. Liamputtong (Ed.), *Social determinants of health: Individuals, society and healthcare*. Melbourne, AU: Oxford University Press.

Mertens, D. M., Bazeley, P., Bowleg, L., Fielding, N., Maxwell, J., Molina-Azorín, J. F., & Niglas, K. (2016). *The future of mixed methods: a five year projection to 2020*. Mixed Methods International Research Association. Retrieved from www.mmira.org

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Richards, L. (2014). *Handling qualitative data* (3rd ed.). London: Sage.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Williams, M. L., & Burnap, P. (2015). Crime sensing with social media. *Methods News: Newsletter from the National Centre for Research Methods*.

Yin, R. K. (2014). *Case study research: design and methods* (5th ed.). Thousand Oaks, CA: Sage.