Guidance in the World of Computer-Assisted Qualitative Data Analysis Software (CAQDAS) Programs

Áine M. Humble

Review Essay:


Abstract: This review discusses Christina SILVER and Ann LEWINS' book, "Using Software in Qualitative Research: A Step-by-Step Guide" (2nd ed.). This book is an impressive undertaking, with online supplemental material in the form of three data sets consisting of many different types of data, detailed instructions for seven CAQDAS (Computer-Assisted Qualitative Data Analysis Software) programs, and full-color reproductions of illustrations from the book. The 14 chapters in the book cover a wide range of analysis issues when working with software programs, and the authors encourage critical use of such tools. Readers will benefit from engaging with the online supplemental tools.

Table of Contents

1. Introduction
2. Situating the Book
3. Overview of the Chapters
4. My Active Engagement With the Book
5. Reflections
6. Conclusion

References

Author Citation

1. Introduction

"Using Software in Qualitative Research: A Step-by-Step Guide" is the second edition of a previous version of this book, written by LEWINS and SILVER in 2007. SILVER and LEWINS take the reader through various steps of using Computer-Assisted Qualitative Data Analysis Software (CAQDAS) programs, to "enable ambitious yet secure use of any CAQDAS package and the moulding of its functions to your needs, while also adding to your awareness of what other tools work well for particular contexts [and] inspire you to explore your data to greater depths, to experiment with software tools and to develop systematic and creative ways of conducting robust and well-evidenced analysis" (p.8). [1]
A substantial amount of online supplemental information is provided at the [companion website](http://www.qualitative-research.net/), all of which can be downloaded. This information consists of three sections: 1. three sample datasets (or "case studies," as they are referred to throughout the book); 2. "step-by-step software guides" for seven CAQDAS programs (ATLAS.ti, HyperRESEARCH, MAXQDA, NVivo, QDA Miner, and Transana), which were created in collaboration with the developers of these programs; and 3. reproductions of illustrations from the book in full-color. The case studies and detailed instructions can be used to explore exercises included at the end of Chapters 4 to 13. [2]

### 2. Situating the Book

CAQDAS programs have existed since the early 1980s (KELLE, 1995; MANGABEIRA, 1995), and it is estimated that there were 15 different programs by 1993 (MANGABEIRA, 1995). Early programs focused on code and retrieve functions (KELLE, 1995), but many new features emerged over time in response to user feedback as well as pragmatic and methodological considerations (RICHARDS, 1999). Different categorizations of CAQDAS programs exist. For example, LEE and ESTERHUIZEN (2000) compared code-and-retrieve programs, code-based theory-building programs, and conceptual network builders; and MacMILLAN and McLACHLAN (1999) differentiated between text retriever, code and retrieve, and theory building programs. Some researchers have argued, however, that program updates have blurred the distinctions originally conceptualized between different programs—particularly between code and retrieve programs and theory building programs (LEWINS & SILVER, 2007, 2009). [3]

Researchers have suggested that the use of CAQDAS has been on the rise and that it is now commonly used (e.g., RICHARDS & MORSE, 2013). Although a reasonable assertion, little empirical research has actually been carried out on this topic. In my own research, I found that family studies journal articles reporting the use of a CAQDAS program increased from 19.4% during the period 2001–2005 to 26.0% during the period 2006–2010 (HUMBLE, 2012), but I noted that this could be an underestimation, as many family studies scholars may use a CAQDAS program but not mention it in their methodology section, or they may publish their work in venues other than journals. CAQDAS use may be very different in other scholarly fields, of course. Anecdotally, it does appear, however, that researchers are increasingly using CAQDAS programs, with graduate students and new faculty members possibly driving the interest (DI GREGORIO & DAVIDSON, 2008). Yet critical discussions regarding how these programs are used in qualitative research or how they influence data analysis are lacking in qualitative methods literature (RICHARDS & MORSE, 2013). With the increasingly complex tools available in such programs, and with the likely increasing numbers of researchers using them, thoughtful discussions are imperative. Researchers need information in order to critically and carefully use such programs and their various features. [4]
The field has been slow to respond to this call to action, but a core group of interested researchers have published on this topic. Interest first appeared in the late 1980s, with the first ever conference on qualitative software held that year in Surrey, England. Literature began appearing in the 1990s. In 2014, a 25-year anniversary conference was held in Surrey. [5]

A few books and journal articles (as well as web resources) exist on this topic. KELLE’S (1995) edited book discussed methodological and theoretical issues related to CAQDAS, but only occasionally mentioned actual software programs. HAHN’S (1998) book, which was written for individuals who did not want to purchase "expensive specialized software packages" (back cover), discussed practical issues related to using everyday software programs such as Microsoft Word and Excel for data analysis. BAZELEY (2007) published a book focused specifically on data analysis carried out with NVivo. DI GREGORIO and DAVIDSON’S (2008) book concentrated on methodological issues, and it included eight case studies that used four different CAQDAS programs. [6]

Occasional articles can be found in discipline-specific journals or books (e.g., HUMBLE, 2012; MATHESON, 2005). FQS has published articles on this topic, including two special issues. The first special issue, edited by GIBBS, FRIESE, and MANGABEIRA (2002), focused on using technology in qualitative research. The second special issue described the "KWALON Experiment" (edited by EVERS, MRUCK, SILVER & PEETERS, 2011), an interesting project that involved five different software developers analyzing the same dataset (guided by the same research question) and then comparing their results. [7]

General books about qualitative data analysis make reference to software, such as CORBIN and STRAUSS’S "Basics of Qualitative Research," now in its third edition (2008), which has used MAXQDA software for its examples. RICHARDS supplemented the second (2010) and third (2015) editions of her book, "Handling Qualitative Data," with an online companion site that involved a discussion of CAQDAS as well as the presentation of ten research projects, with each including comments about their software use. As web-based resources have developed, online companion sites to books have increasingly made sense, particularly for the CAQDAS field, which advances so quickly. In their third edition (2013) of "Readme First for a User’s Guide to Qualitative Methods," RICHARDS and MORSE referred readers to the companion site developed for RICHARDS’ book. [8]

On that note, one very important source of information on CAQDAS is a website called the CAQDAS Networking Project, which has existed since 1994. LEWINS managed this website from 1994 to 2010; SILVER became involved in the project in 1998 and has managed it since 2010. This website provides information about CAQDAS training, reviews of software packages, and articles on topics such as choosing a software package, working in teams, and analyzing survey and visual data. [9]

LEWINS and SILVER are the authors of this book being reviewed, which was first published in 2007 in response to a need for practical support and "step-by-step"
guidance in using software programs. The first edition consisted of detailed instructions for various CAQDAS software programs. However, because such software changes at a rapid pace, hard copies of instructions can become quickly outdated. Thus, this second version has moved those detailed instructions to a companion website. This edition also provides more content on "methodological contextualisation" (p.4), visual methods of data analysis, and mixed methods approaches, as well as includes more software packages in its discussions and examples. [10]

3. Overview of the Chapters

This book consists of 13 chapters, which are presented in a logical order, but the authors point out that just because they are in a particular order does not mean that a researcher's analysis proceeds in a linear fashion. For example, working with coding schemes does not have to occur after coding. Exercises are presented at the end of Chapters 4 through 13, and PDF files for instructions corresponding with each of these chapters can be downloaded from the companion website. Here is an example of a handout for MAXQDA from Chapter 5. Readers can download the data files from the companion website or use their own data—the exercises are presented generally enough so that this is possible. Chapters also contain helpful tables and figures (typically, screenshots from various software programs), as well as reference the case studies. [11]

The first chapter introduces the reader to CAQDAS and discusses general issues related to qualitative analysis, such as planning and managing a project, coding, writing memos, retrieving segments, and creating visual representations of data. Six key tasks for setting up a software project are discussed. The non-linear and flexible nature of qualitative data analysis is discussed, with helpful distinctions made between factors such as code-based and non-code-based approaches to analysis and "horizontal cuts" (examining codes across different transcripts) versus "vertical cuts" (examining codes in detail for one case) through data. The chapter concludes with a brief description of seven strategies (discourse analysis, narrative inquiry, framework analysis, grounded theory, thematic analysis, mixed methods research, and visual analysis) that can be supported with CAQDAS. The authors point out that this list is not exhaustive. [12]

In the first part of Chapter 2, the case studies are presented, including suggestions for possible analysis. These case studies are:

1. Case Study A: A "theory-driven abductive" study about young people's perceptions of how they learn about sex and relationships (files are literature, interview transcripts, responses to vignettes and photographs, and demographic data).

2. Case Study B: A "theory-informed abductive" study about the global financial downturn (files are focus group discussions, surveys, and media releases).
3. Case Study C: A "theory-building inductive" study about Coca-Cola commercials and gender over a 50-year period (files are commercials in mp4 video format and company documents). [13]

The second half of this chapter describes general processes of integrating, organizing, exploring, reflecting, and interrogating qualitative data using CAQDAS, such as coding, grouping pieces of information together, hyperlinking, making annotations and memos, and using tools to identify patterns. [14]

Chapter 3 summarizes the eight software programs used in the book. The authors point out that most of the summaries were written in collaboration with the developers of those programs and that the information was accurate as of August 2013. A table at the end of the chapter presents additional resources and less commonly used software applications used for data processing (such as transcribing), analysis, mapping, and capturing or processing data in multimedia format. Web addresses are provided for each resource. [15]

How data needs to be prepared for importing into a software program is discussed in Chapter 4. The authors distinguish between data that is fully integrated into a project versus data that is incorporated or simply referenced. Many file formats are able to be imported into programs, such as textual data (e.g., rich text format, doc, or docx files); PDF files, images; and audio and video files. Quantitative data can also be imported into software programs for studies involving mixed methods. This chapter points out the importance of knowing what special formatting is required in a software program for both qualitative and quantitative data before importing the data in. Helpful points such as "create an efficient naming protocol for files," "avoid transcribing into tables," and "consider whether you will need to edit [...] after import" (p.89) are offered. Multimedia preparation of data is also discussed, such as the ability to synchronize audio or video files to transcripts. [16]

Chapter 5 discusses early steps in the development of a project ("project" meaning the software file containing the whole analysis—the imported files, codes, memos, and so on), such as deciding on a procedure to keep track of analysis (logs, memos, audit trails), naming and backup protocols, and considering how a project will allow for growth (the addition of future material, integration of new material, or reorganization of existing material). Considering whether or not and how to integrate a literature review into the program is also discussed. This could be through importing actual PDF files of journal articles and reports, but it could also be through importing bibliographic information from bibliographic programs such as EndNote and Zotero. [17]

Chapter 6 describes ways of exploring the data that involve marking and annotating data (e.g., applying a comment to a piece of text), searching through the content (e.g., for words or phrases), and linking points in data. These exploration alternatives are presented as ways to become familiar with one's data, and include procedures for becoming familiar with the data other than

© 2015 FQS http://www.qualitative-research.net/
through coding (which many people have argued is overly emphasized in CAQDAS programs). [18]

Coding is the focus of Chapter 7. Coding terminology is introduced, and the authors distinguish between the following approaches to coding: 1. inductive, 2. deductive, 3. theoretical, and 4. question-based. A discussion of inductive and deductive coding strategies is included, with a note that a dichotomous view of this is not necessarily helpful. The combination of induction and deduction is called abduction. Flexibility in coding is important, and it is emphasized that coding is not analysis "in and of itself" (p.181). [19]

Chapter 8 is about "basic" retrieval of coded data, which is important not only for comparison of coding but also for giving a snapshot of where a researcher currently is in their data analysis, helping to identify areas for future analysis, and pointing out where data may need to be recoded. These kinds of decisions, of course, are based on the researcher's interpretation of the data that the software program provides the means of reviewing. As many others have indicated (e.g., HUMBLE, 2012; RICHARDS, 2015), the software program does not do the analysis for the researcher; the researcher is still in charge of the interpretation of the data. Different types of retrieval are discussed: quantitative overviews, horizontal cuts, vertical cuts, and filtering strategies that allow for comparison of subgroups (through short-cut grouping such "sets" or "families"). [20]

Chapter 9 discusses coding schemes, which are how a software program lists the researcher's codes. Some software packages allow for a hierarchical presentation of codes whereas others do not, and the authors point out that programs allowing hierarchical coding schemes may have subtle differences (e.g., the higher-level code may or may not contain data coded in the subcodes). Diverse ways of setting up a coding scheme are presented. One interesting example is to have top-level codes called "broad-based codes," "emergent codes," and "sensitizing concepts," and then later to reorganize subcodes from within those categories into different sets of codes (with sets or families). Various factors that influence a coding scheme (such as its relationship with a theoretical framework) as well as guiding principles for developing coding schemes are also included. The chapter concludes with some points about maintaining a coding scheme. [21]

Chapter 10 is a key chapter discussing the writing features of software programs, and it brings together many points from previous chapters. The authors describe the importance of writing within the programs and the varying purposes of doing so, and they outline different ways of doing so (such as code definitions and analytic memos, process memos [e.g., logbooks], and summaries). Many tips are provided with regard to tasks such as creating and naming memos, linking memos to data, and coding and searching the researcher's own writing. [22]

The focus of Chapter 11 is on mapping ideas (creating visual representations of data). Reasons for engaging in mapping are discussed. For example, one researcher might use mapping to create a theoretical model, whereas another
person might simply use it to review their work or explore some hunches. Practical suggestions for using mapping functions are presented. General mapping functionalities in CAQDAS packages are described and some key differences between packages are noted (such as mentioning which programs allow for multiple map layers). [23]

Chapter 12 discusses how data can be organized by "known characteristics," which is particularly helpful when working with large data sets. Such organization allows the researcher to focus on subsets of data within a larger dataset, such as on the basis of demographic characteristics (e.g., gender, geographical location) or characteristics that have arisen out of the data analysis. Organization may also occur within a document, such as speaker sections within a focus group. Other circumstances that might benefit from data organization are those in which change over time is being tracked, or multiple data files are being combined to create one unit of analysis. Basic principles for organizing data are presented. [24]

The last main chapter (Chapter 13) in the book focuses on interrogating datasets, which refers to strategies "oriented around identifying patterns, relationships and anomalies, comparing cases and subsets, and testing theories and quality more generally" (p.303). Interrogating a dataset occurs in many ways, not just through the formal theory-building (querying) features of programs. SILVER and LEWINS use the term "querying" in a broader sense, seeing it as a "key means by which work moves from simple code-and-retrieve to more in-depth analysis" (p.303). Although this topic is the focus of the final chapter, the authors point out that it can occur before coding is completed. Various reasons for engaging in interrogation are discussed, as well as ways of interrogating data presented (e.g., filtering code stripes in the margin so that only certain code stripes are viewed, qualitative cross-tabulations, and coding queries [those using Boolean operators]). The chapter concludes with a discussion of how interrogations may be presented visually in software programs through tables, matrices, charts, and graphs. [25]

The conclusion chapter (Chapter 14) revisits key points raised throughout the book, such as the need to prepare for using a software program, how software can bring researchers closer to their data, and the importance of using software tools critically and carefully (particularly those regarding automatic features, quantification of data, and mixed methods). [26]

4. My Active Engagement With the Book

SILVER and LEWINS' book and its accompanying website are meant to create an interactive experience for readers. This makes sense. A researcher must explore a software program in order to understand its capabilities; reading about it is simply not enough. Thus, in my review of this book, I also downloaded one of the case studies (Case Study B), imported its data into MAXQDA, and then followed some of the exercises in Chapters 4, 5, and 6, including looking at the detailed instructions for those chapters (e.g., https://study.sagepub.com/sites/default/files/MAXQDA_Silver

© 2015 FQS http://www.qualitative-research.net/
Here is a list of some of the tasks I carried out as part of my review.

- Chapter 4: I reviewed the raw data (and made some edits and formatting changes to it) before importing it into the program.
- Chapter 5: I created document groups, and then imported the data in, including survey data that was in an excel file. I autocoded this survey data upon import, which, as noted by the authors, necessitated some pre-formatting of the Excel file before it could be properly imported in. I created three empty sets. I also created memos within documents as well as for document groups, document texts, and codes.
- Chapter 6: As part of my review of the text (becoming more familiar with it before engaging in coding, as SILVER and LEWINS suggest), I created an in-document link and made a link between two pieces of text. I conducted a search for "anx*" (to catch the words "anxiety" or "anxious"), which resulted in two hits.

5. Reflections

This publication is an impressive and exhaustive undertaking (with a book, three data sets consisting of many different types of data, and detailed instructions for seven CAQDAS programs), particularly given the depth, breadth, and ever-changing nature of this field. The authors are to be commended for tackling this very challenging topic. I have been facilitating workshops on how to use MAXQDA in Canada and the U.S. for several years and I am very familiar with this software package, and I was delighted to learn some new things about it as a result of conducting this review. I was also introduced to some new concepts and suggestions for ways of coding that I had never thought of before (e.g., initially setting up code categories for "emergent codes," "broad-brush codes," and "sensitizing concepts," and then combining groups of those codes into various sets).

A definite strength of the book is the inclusion of examples from many different programs. Although ATLAS.ti, MAXQDA, and NVivo are the most commonly used programs (in my opinion, and I think others would agree), it is very valuable to see examples of other programs and what they can offer to the qualitative researcher. Readers should be aware, however, that at the companion website, not all of the chapters include step-by-step instructions for all seven software programs (for example, ATLAS.ti, MAXQDA, and NVivo do not have guides for Chapter 4; Dedoose excludes Chapter 11; and Transana excludes Chapters 4, 6, 9, 11, and 12).

I did not complete later chapter exercises that would have required me to first code the sample data, as I did not have time to do that. Having a full copy of the dataset with codes already applied would have encouraged me to complete...
exercises in later chapters (e.g., Chapters 11 to 13), but I do understand why the authors did not include this. However, it could be offered as an option at the companion website. [31]

There is room for some revisions of the companion website, which will be easy for the authors and Sage to do. I could not find any information about teamwork even though it was mentioned that there was a separate section on the website about this topic (an important topic that the authors note is not included in the book). I worked with Case Study B (the financial data), and in reading the step-by-step instructions for Chapter 5, I realized that a bibliography document could also be imported into the program, but I did not see this included in the zip file containing all the files. Additionally, before importing the documents into MAXQDA, I noticed that the transcripts contained some spelling errors and formatting inconsistencies in headings. I do not know if this was deliberate or not; it is possible that the authors may have wanted people to review the transcripts before importing (as an important part of preparing one's data) and to determine through trial and error whether they can edit a document once it is imported into a program. Again, this may be part of a bigger plan to ensure that readers review and standardize their files before importing them into a software program (in which case, bravo, SILVER and LEWINS, how clever of you!). My experience with these inconsistencies certainly got me thinking more about these kinds of issues, which was beneficial in and of itself. [32]

As noted, I am very experienced with MAXQDA software and I have also taught myself to use older versions (from five years ago) of ATLAS.ti and NVivo, but I found some of the language overwhelming and a bit confusing at times. For someone with less experience, it is key that they use the datasets and the step-by-step instructions to help them understand the concepts. RICHARDS’ (2015) easy-to-read book on "handling qualitative data" could be an excellent complement to this resource. [33]

CAQDAS software programs have developed tremendously in the past decade, and many new features have emerged in these programs. In my own work, both written (HUMBLE, 2012) and consulting, I have encouraged researchers to be critical about the "bells and whistles" in various programs. Just because these bells and whistles exist does not mean that they need to be used, or used in every qualitative or mixed-methods study conducted. SILVER and LEWINS do a very good job of similarly cautioning their readers to "never do something just because it is possible" (p.34), and of pointing out that an analysis can be taken away from the computer when necessary (e.g., reading hard copies of transcripts and making notations on them). [34]

A couple of minor points: I would have liked to have seen some ethical considerations noted with regard to synchronizing transcription to audio and video files (potential loss of confidentiality). Additionally, a type of memo that could be mentioned in Chapter 10 is one that keeps track of how the researcher uses the software program. Researchers often provide few details about how they use the CAQDAS programs. Keeping track of this systematically in a memo would be a
good step toward documenting their use, and could be included as part of an audit trail. [35]

Finally, as I progressed through the book, I began to envision a university-level course on CAQDAS being offered, which I think would be a very useful experience. The text and its companion website could be used as the central reading for the course (the step-by step handouts contain a note that if someone wants to use the handouts for teaching purposes, they should contact info@qdaservices.com). Such a course could be taken after completing a qualitative methods course, and students could use their own thesis or dissertation data to explore the exercises provided in the book. Those who do not have data yet could use the datasets provided by the authors. Although it would make sense to use just one software program in the course (I cannot imagine a professor trying to teach more than one software program in such a course), students would come away with an in-depth knowledge of what that particular software program was capable of, which could be transferred to other contexts (software programs). [36]

6. Conclusion

This book is a valuable contribution to the CAQDAS field, providing a critical and in-depth discussion of what CAQDAS programs have to offer, offering examples from many different software programs and including online supplemental material. Researchers who choose to both immerse themselves in the book and take advantage of the online resources and related exercises will be richly rewarded. The challenge may be finding the time to do all of it, but those who only focus on certain parts of the book will also benefit. I definitely see value in using this book to structure a university-level course focusing specifically on CAQDAS. Such a prescribed structure will ensure that the readers really engage with the book as it is intended. [37]

References


Author

Áine HUMBLE is associate professor of family studies and gerontology at Mount Saint Vincent University, Canada. Her research interests focus on 1. family rituals, 2. women and healthy aging, 3. same-sex couples, 4. qualitative research, and 5. CAQDAS. She is currently working on a study of older LGBT (lesbian, gay, bisexual, and transgender) individuals and end-of-life preparation. She has been providing MAXQDA training since 2009 through her consulting company, Fada Research Consulting.

Contact:

Dr. Áine M. Humble
Department of Family Studies and Gerontology
Mount Saint Vincent University
166 Bedford Highway, Halifax, Nova Scotia, Canada, B3M 2J6
Tel.: ++1 902-457-6109
E-mail: aine.humble@msvu.ca, fadaconsulting@gmail.com
URL: http://www.msvu.ca/fsgn/faculty/humble

Citation