Using the SAS® System and QSR NUD*IST®
for qualitative data analysis (QDA)

Robert G. Stewart, East Tennessee State University, Johnson City, TN

Abstract

With the advent of mixed-method inquiry (i.e., the combining of procedures meant to gather and analyze different information), quantitatively imbued SAS users will eventually encounter (if not already) some raw (uncoded) form of qualitative data (e.g., texts, images, sounds). In this paper, the requirements and capabilities of the SAS System and QSR NUD*IST 4 (Nonnumerical Unstructured Data Indexing Searching and Theorizing by Qualitative Solutions and Research [called "N4" for short]) regarding such data are explored. The overriding objective is to assist SAS users at all levels in making effective decisions regarding qualitative data analysis (QDA). In doing so, the following questions are addressed: (a) What forms of data might one encounter? (b) What approaches to analysis might one use? (c) What can one do with SAS? (d) What can one do with N4? (e) What other QDA programs are available? Free-flowing texts produced by responses to open-ended questions (a common form of uncoded qualitative data) are emphasized in discussing the analysis requirements and capabilities of SAS and N4.

Introduction

"This is an era of methodological pluralism in applied social science, including the field of evaluation. Multiple frameworks for inquiry abound. Interpretivist, postpositivist, activist, literary, feminist, and critical frameworks, among others, compete for our attention and allegiance" (Greene & Caracelli, 1997, p. 5). And so begins a recent treatise on mixed-method inquiry (i.e., "inquiry that intentionally combines different methods. . . meant to gather different kinds of information. . . . to develop important knowledge claims that respect a wider range of interests and perspectives; Greene & Caracelli, 1997, p. 6). With the advent of mixed-method inquiry, feelings of elation or consternation (or both) may be aroused in either quantitatively or qualitatively imbued researchers. Elation in that one may have opportunities to promote and practice hard--won competencies. Consternation in that one may have to integrate into or (perhaps even worse) master different (perhaps even diametric) forms of data and approaches to analysis. Indeed, those from the quantitative camp (which albeit includes most SAS users) are bound to encounter (if not already) a raw (uncoded) form of qualitative data (e.g., texts, images, sounds). How then should one respond?

A basic mixed-method design most are likely to find, is a survey, regardless of format (e.g., mail, telephone, face-to-face), that solicits responses to closed and open-ended questions. In such a case, both scaled (nominal, ordinal, interval, ratio) and uncoded (raw) data may be produced. For this paper, the focus is on uncoded qualitative data since it is assumed (I hope correctly) that SAS users are already familiar with scaled data. It should be noted that the term unscaled is avoided, because I am not satisfied that texts, images, and sounds are not (somehow) scaled while in an uncoded (raw) form.

To one's disappointment (or delight?), treatises on the analysis of mixed-method data at either the technical (method) or philosophical (paradigm) level are lacking in the literature (but see Caracelli & Greene, 1993, for some suggestions). Likewise, while primers are available on the integration of SAS with other statistical packages (e.g., SPSS, Minitab; Matthews, 1996) works for integrating quantitative and qualitative analysis software (say e.g., SAS and N4 in analyzing the survey above) are scarce. Indeed, the following statement by Weitzman (2000), is the only relevant example that I could find:

> If your data, and/or your analyses, include the possibility of number crunching, look to see whether the program will count things and/or whether it can share information with quantitative analysis programs such as SPSS or SAS. Think carefully about what kind of quantitative

analysis you'll be doing, and make sure the program you are thinking about can arrange the data appropriately. Consider, too, whether the program can link qualitative and quantitative data in a meaningful way (in terms of the analytic approach you are taking). For example, do you need to be able to select subsets of you qualitative data based on quantitative scores or demographics? Or do you need to use your qualitative coding to generate scaled variables for statistical analysis in SPSS [or SAS]? Or do you want to be able to generate word or code frequency tables for statistical analysis? (pp. 814-815)

While informative, seemingly important questions from both quantitative and qualitative camps go unanswered (e.g., What kinds of statistical analyses are theoretically appropriate for frequency tables derived from texts, images, and sounds?). Faced with such expositional shortfalls, my objective here is to assist SAS users at all levels in making effective decisions regarding qualitative data analysis (QDA). In doing so, the following questions are addressed: (a) What forms of data might one encounter? (b) What approaches to analysis might one use? (c) What can one do with SAS? (d) What can one do with N4? (e) What other QDA programs are available?

Forms of Data

One could think of qualitative data as a synonymous classification for categorical response data (i.e., variables measured on nominal and possibly ordinal scales; Agresti, 1990, p. 4, para. 1.1.4). However, I find that the concept of categories (either ordered or unordered) does not completely reflect the characteristics of the raw (uncoded) qualitative data forms of interest here (and elsewhere; Denzin & Lincoln, 2000) which are texts (e.g., responses to open-ended questions, documents), images (e.g., videos, photographs), and sounds (e.g., music, noise). Of these, free-flowing texts produced by responses to open-ended questions (a common form of uncoded qualitative data; Ryan & Bernard, 2000) are emphasized in discussing the analysis requirements and capabilities of SAS and N4. For more information on uncoded qualitative data, readers should consult Bauer and Gaskell (in press, for texts, images, and sounds), Harper (2000, for images), and Silverman (2000, for texts).

For the most part, textual data stem from two inquiry traditions: linguistic and sociological. The former "treats text as an object of analysis" while the latter "treats text as a window into human experience" (Ryan & Bernard, 2000, p. 769). Of these, I will focus on the sociological tradition in which two kinds of written texts are recognized: (a) "words or phrases generated by techniques for systematic elicitation" and (b) "free-flowing texts, such as narratives, discourse, and responses to open-ended interview questions" (Ryan & Bernard, 2000, pp. 769-770). It should be noted that open-ended questions can also produce words or phrases (e.g., free lists; Ryan & Bernard, 2000).

A further (and useful) delineation of textual data is possible: the magnitude (for lack of a better term) of description. In the case of description, two classifications are recognized: thin and thick (Creswell, 1998). According to Denzin (1989b), a thick description "presents detail, context, emotion, and the webs of social relationships . . . [and] evokes emotionality and self-feelings. . . . The voices, feelings, actions, and meanings of interacting individuals are heard" (p. 83 as cited in Creswell, 1998, p. 184) (insertion unknown). To contrast, he provides the following example of a thin description "I had trouble learning to play the piano keyboard" (Denzin, 1989b, p. 85 as cited in Creswell, 1998, p. 184).

Approaches to Analysis

As is the case with quantitative data, qualitative data also benefits from reduction techniques (Ryan & Bernard, 2000). Accordingly, many approaches to analysis have been developed. Unfortunately, only a cursory treatment of these techniques is possible here. While terminology can differ among authors, in general, analysis approaches can be grouped into one of four categories (Ryan & Bernard, 2000): words and phrases, cultural domains, free-flowing texts, and units of thought. An extended classification now follows:

words and phrases
> free lists
> paired comparisons
> pile sorts
> triad tests
> frame substitution

cultural domains
> componential analysis
> taxonomies
> mental maps

free-flowing texts
> key-words-in-context (KWIC)
> word counts
> structural analysis
> cognitive maps

units of thought
> grounded theory
> schema analysis
> visual displays
> content analysis
> content dictionaries
> analytic induction and Boolean tests
> ethnographic decision models

As one is perhaps experiencing either elation or consternation, a caveat is worth noting: no one method of qualitative data analysis can satisfy all inquires (Ryan & Bernard, 2000). So perhaps elation suffers (I hope minimally).

At this point, it would be poor form to discuss qualitative data forms and analysis approaches without mentioning inquiry paradigms. Indeed, just as there are different quantitative paradigms (e.g., Fisherian, Bayesian statistics), so to exist qualitative ones (e.g., biography, phenomenology, grounded theory, ethnography, case study). Moreover, mixed method inquiry boasts its own paradigmatic stances (e.g., purist, pragmatic, dialectical; Greene & Caracelli, 1997, pp. 8-9). However, a detailed exposition, while desirable, is not possible here. Indeed, I assert that data forms and analysis approaches are more relevant for SAS users than details of one or more inquiry paradigms--at least initially. For more information on qualitative paradigms, readers should consult Creswell's (1998) lucid text.

SAS for QDA

Requirements. When faced with analyses of mixed-method data (perhaps for the first time), one is likely to make the following assumption: numbers should be analyzed quantitatively while texts should be analyzed qualitatively. However, other combinations are possible as Bernard (1996) points out: "qualitative data (text) and quantitative data (numbers) can be analyzed by quantitative and qualitative methods" (p. 10). It may help to think of each combination as a cell in a two-way table. For example, statisticians perform a qualitative analysis on quantitative data when interpreting the output from statistical tests. One is encouraged to ponder: "To what extent do quantitative and qualitative analysis software (e.g., SAS and N4) follow this logic?" For example, "Can SAS be used to analyze free-flowing texts produced by open-ended questions?" The answer is yes, provided that the data are transformed. Bernard (1996) describes the required transformation as follows:

> Strictly speaking there is no such thing as a quantitative analysis of qualitative data. The
> qualitative data (artifacts, speeches, ethnographies, TV ads) have to be turned first into a matrix,

where the rows are units of analysis (artifacts, speeches, cultures, TV ads), the columns are variables, and the cells are values for each unit of analysis on each variable. (p. 10)

Capabilities. Assuming that textual data have been transformed into rows, columns, and values, "What can one do with SAS?" In short, many analytic options are possible. For example, (a) frequencies, (b) correlations, (c) discriminant analyses, (d) two-way contingency tables, and (e) cluster analyses are all supported by SAS. It should be noted that these analytic options are particularly useful for content analyses (Krippendorff, 1980). In matching analytic options with SAS codes, readers should consult applicable documentation (e.g., Stokes, Davis, & Koch, 1995).

Just as coding is essential to qualitative data analysis, so to are calculations for inter-coder agreement and reliability (Denzin & Lincoln, 2000). Indeed, many measures (e.g., kappa coefficient; Cohen, 1960) and designs (single vs. multiple coders, overlapping vs. nonoverlapping observations) are available. In most of these cases, SAS can be used to calculate the needed estimates (for an example of a multiple rater-criteria design, see Spector 1993, pp. 109-113).

SAS can be used exclusively to analyze free-flowing texts. However, based on my experience it is only parsimonious when (a) the number of actual observations and open-ended questions is small (say, 50 informants, 4 questions), (c) the data constitute very thin descriptions (e.g., word lists), and (d) verbatim transcripts are not desired (perhaps to ensure anonymity). Indeed, anything beyond these points, and one should consider a computer-aided qualitative data analysis (CAQDA) option (e.g., N4).

N4 for QDA

Requirements. With regard to CAQDA, Weitzman (2000, p. 805) makes an important point through a somewhat faulty contrast:

Simply put, software [e.g., N4] can provide tools to help you analyze qualitative data, but it cannot do the *analysis* for you, not in the same *sense* in which a statistical package like SPSS or SAS can do, say, multiple regression. [sic, the *sense* is equivalent in that both quantitative and qualitative data software do *analyses* which require programming and interpretation by the user] (emphasis current).

At any rate, the main point is that to use analysis software effectively, one must have achieved a level of competency. With regard to N4, this means having the ability to code texts (or in a word, coding). Ryan & Bernard (2000) outline the operational requirements for coding as follows: (a) selecting units of analysis within the texts, (b) finding themes, (c) organizing a codebook, (d) assigning codes to units of text, and (e) linking themes to one another based on a conceptual model. One should note that prior to coding, texts must be transcribed (if audio-taped), formatted with a work processor, and then imported into N4. The process can consumer either one's time (if done by self) or money (if contracted out). For more information on inductive coding, readers should consult the cogent work by Boyatzis (1998).

Capabilities. Assuming that the data have been transcribed, formatted, and imported, "What can one do with N4?" In short, more than one probably wants to do. Indeed, seven core capabilities are available. These are (a) adding and deleting codes, (b) creating and coding annotations, (c) memoing, (d) creating and editing an index system, (e) searching text for key terms, (f) querying the index system, and (g) importing or exporting tables. Moreover, 17 search operators (e.g., the union command) are available for the index system (see QSR NUD*IST 4 User Guide, 1997, p. 136).

Given a substantial number of observations and open-ended questions, data that constitutes thick description, and a requirement for verbatim transcripts, the above capabilities become parsimonious (if not indispensable). Indeed, it is these very characteristics that push a researcher to be more qualitative (e.g., thoughts as units of analysis) rather than quantitative (e.g., counts of words).

At this point, for those experiencing consternation, it should be noted that a comprehensive tutorial is

provided with the N4 software. For more information, readers should consult the lucid primer by Gahan and Hannibal (1999) and the QSR NUD*IST 4 User Guide (1997, for those short on reading time, begin with p. 39). One should also note that version 5 of NUD*IST (called *"Nvivo"*) is now available.

Other QDA Programs

Undoubtedly, other computer-aided QDA tools are available. However, given the space remaining, only evaluative resources (not the individual packages) can be discussed. For a thorough review of available software, readers should consult Weitzman and Miles (1994). Also, Miles and Huberman (1993, p. 316) provide a comparative analysis of 22 software packages. Moreover, Weitzman's (2000) chapter–*Software and Qualitative Research*–is essential reading–at least in my opinion. It assists the reader in selecting software by providing an exhaustive set of criteria (e.g., skill level of user, project duration, required analyses)*.*

Conclusion

The requirements and capabilities of the SAS System and QSR NUD*IST 4 regarding the analysis of uncoded qualitative data (viz., free-flowing texts produced by open-ended questions) have been explored. It is hoped that others are inspired to further explore the integration of quantitative and qualitative analysis software with uncoded qualitative data forms (e.g., images and sounds). Moreover, It is hoped that feelings of consternation among quantitatively imbued SAS users when contemplating mixed-method data have been placated–at least somewhat. In closing, I find this assertion by Greene & Caracelli (1997)   appropriate:

> In this troubled era, with social problems of ever-increasing complexity and intractability, multiple ways of knowing and acting are surely needed" (p. 15).

References

Agresti, A. (1990). Categorical data analysis. New York: John Wiley & Sons.

Bauer, M. W., & Gaskell, G. (in press). Qualitative research with text, image and sound: A practical handbook. Thousand Oaks, CA: Sage.

Bernard, R. (1996). Qualitative data, quantitative analysis. Cultural Anthropology Methods Journal, 9-11.

Boyatzis, R. E. (1998). Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, CA: Sage.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 15(2), 195-207.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-48.

Creswell, J. W. (1998). Qualitative inquiry and research design: Choosing among five traditions. Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (2000). Handbook of qualitative research (2nd ed.). Thousand Oaks, CA: Sage.

Gahan, C., & Hannibal, M. (1999). Doing qualitative research using QSR NUD*IST. Thousand Oaks, CA: Sage.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method

evaluation. In J. C. Greene & V. J. Caracelli (Issue Eds.), G. T. Henry & J. C. Greene (Series Eds.), Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (pp. 5-17). New Directions for Evaluation, 18(2 , Serial No. 74). San Francisco: Jossey-Bass.

Harper, D. (2000). Reimagining visual methods: Galileo to *Neuromancer*. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed.). (pp. 717-732). Thousand Oaks, CA: Sage.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage.

Matthews, G. (1996, October). Integrating the SAS® system with other statistical applications. Paper presented at the annual meeting of the Southeast SAS Users Group, Atlanta, GA.

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

QSR NUD*IST 4 User Guide (2nd ed.). (1997).
Victoria, Australia: Qualitative Solutions and Research Pty Ltd.

Ryan, G. W., & Bernard, R. (2000). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed.). (pp. 769-802). Thousand Oaks, CA: Sage.

Silverman, D. (2000). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed.). (pp. 821-834). Thousand Oaks, CA: Sage.

Spector, P. E. (1993). SAS® programming for researchers and social scientists. Newbury Park, CA: Sage.

Stokes, M. E., Davis, C. S., & Koch, G. G. (1995). Categorical data analysis using the SAS System. Cary, NC: SAS Institute, Inc.

Weitzman, E. A. (2000). Software and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed.). (pp. 803-820). Thousand Oaks, CA: Sage.

Weitzman, E. A., & Miles, M. B. (1994). Computer programs for qualitative data analysis. Thousand Oaks, CA: Sage.

Acknowledgments

Author

Please send your comments or requests to:

stewart_gris@hotmail.com