The *bricoleur* with a computer, piecing together qualitative and quantitative data.

Patricia Bazeley

University of Western Sydney Macarthur

Address for correspondence:
Dr Pat Bazeley
University of Western Sydney Macarthur
PO Box 555
Campbelltown NSW 2560
Australia.
email: p.bazeley@uws.edu.au

Published in: *Qualitative Health Research, 9*(2), 279-287.

Abstract

The researcher as *bricoleur* will gather whatever data is at hand, experimenting and exploring to find answers to the questions he or she has set. With computer in hand and new tools available she can readily combine data types, moving beyond complementarity and simple triangulation. Data may be transferred in either or both directions between NUD*IST (a program to assist the analysis of qualitative data) and a spreadsheet or statistical package. Thus analysis and interpretation are enriched, and new ways of thinking about data are laid open. Such techniques inevitably challenge traditional assumptions about particular methods. But perhaps in the final analysis all methods, other than those employed in reductionist, hypothesis testing experiments, are essentially interpretive.

It is more than twenty years since Denzin (1978) developed Webb, Campbell, Schwartz and Sechrest's (1966) metaphor of triangulation for research methodology, suggesting that the use of multiple data sources, methods, investigators and theories contributed to greater reliability and validity of results in social science research. Denzin and Lincoln, in the introduction to their 1994 *Handbook of Qualitative Research*, describe qualitative research as "inherently multimethod" and the qualitative researcher as a "*bricoleur*", piecing together emergent solutions to a puzzle (although Barbour [1998] would take issue with the appropriateness of multiple methods even within a qualitative framework). With advances in computer software it is now possible to integrate elements of quantitative with qualitative data in a way not previously feasible. One can *fuse* qualitative and quantitative data within a single analysis and so to be a *bricoleur* in an even wider sense, thus enhancing one's ability to understand the issues at hand and solve the puzzles of research and practice.

Quantitative and qualitative methods were initially associated with different disciplinary traditions such that *separation* was generally maintained between them. Although it is now widely recognized that there has always been a level of mixing (e.g. interviews undertaken in order to design an appropriate questionnaire), writing about the *combination* of methods has gained momentum only in the last two decades. Typically, that writing has been about using quantitative and qualitative methods in a complementary or comparative way, either sequentially or in parallel, and most often, with one method taking priority over the other (e.g. Creswell, 1994; Morgan, 1998; Morse, 1991). Mixed method designs were classified into two primary groups by Caracelli and Greene (1997): component designs (such as those just described) and integrated designs, with the latter including (among others) iterative designs, nested designs and holistic designs. I propose that what has now become possible is the *fusion* of qualitative and quantitative methods, where qualitative and quantitative data are used interdependently in the same set of data to satisfy the requirements of these integrated designs, and beyond.

Computers have always been able to deal with numbers as data, so quantitative analysis has, in a sense, had a 'head start' in so far as computers have made it possible to process large volumes of data and/or carry out statistical procedures requiring complex or iterative calculations. Programs purpose built to facilitate analysis of text-based data are a relatively recent addition to the tools of the qualitative researcher – sufficiently new still to be shunned by those 'purists' who see them as reductionist and decontextualizing. Now computer programs are being developed which can make analytical use of both textual and numeric data, with most qualitative analysis programs allowing for at least some numeric output. WinMAX (Kuckartz, 1997) was designed specifically with blending data in mind and allows for import of statistical information into a qualitative database and for integration of coded information back into a statistical program. Also, in mid-997, SPSS (Statistical Package for the Social Sciences) released Text Smart, a module for content analysis of text (although content analysis is still seen and used primarily as a quantitative method). While it has long been possible to incorporate coded numeric data into a NUD*IST database (advanced by the realisation, in 1993, that text search and command language tools could be used to 'automate' this process) and to create numeric output (somewhat more laboriously), the capacity to combine numeric and text data within the

same analysis was greatly enhanced with the advent of NUD*IST rev. 4 (Qualitative Solutions and Research, 1997). In this release, the desire to import and export statistical data (for example, directly from and to SPSS) was specifically addressed with the introduction of table import and export functions – exported data being either information about codes which had been applied to the text, or a matrix built from crosstabulated, coded data. The consequent possibilities for blending qualitative and quantitative data range, therefore, from simple sorting of qualitative comments according to some categorical or scaled criteria, to incorporating the results of qualitative coding in correspondence analysis, logistic regression or other multivariate statistical techniques. From a technical point of view then, blending qualitative and quantitative data is no longer difficult, providing one is reasonably comfortable with computers and with moving between a 'potpourri' of computer programs.

The approach taken to research design and analysis here is a pragmatic one (Datta, 1997; Greene and Caracelli, 1997; Patton, 1989) in which any data which can contribute to an understanding of the issues at hand are seen as worthy of consideration. This means that the stance taken at any particular time could be post-positivist, interpretive or constructivist (Guba & Lincoln, 1994), with epistemologic choices depending on the issues raised by the question and the context and complexity of the analysis rather than on the type of data available (i.e. whether it is numbers or text or a combination of the two).

Fusing data using computers

I routinely use SPSS, Word and Excel with NUD*IST 4 for the task of combining data types, although other statistical, word processing, spreadsheet and database programs could be used equally well. NUD*IST 4 and winMAX are currently the only qualitative analysis programs that readily facilitate combining data types. In this article I focus primarily on the technical side of combining data, illustrating what can be done and how it might be useful from a range of projects I have been involved with (as either chief investigator or consultant). I will then briefly review the issues that arise.

Access to text within a primarily quantitative analysis: The first and simplest illustrations are examples of what might be additionally gained from survey responses when both statistical information and the full text of open-ended responses, examples and comments are readily accessible to the analyst in the same database. Standard treatment of text responses in surveys has traditionally involved simple categorisation and counting, with perhaps categorisation and coding to allow for computer entry and cross tabulation or correlation with other variables (with the Multiple Response function on SPSS being of great benefit here in allowing computations for multifaceted text responses). But the capacity to evaluate the text itself against quantitative variables from the survey has the potential to greatly enrich the interpretation of the data. From the qualitative end, NUD*IST has always encouraged the incorporation of factual data about respondents in the qualitative database so that, for example, questions can be asked about whether there are differences in the nature of the responses given between those of different genders, or ages, or religious affiliations, and so on (Richards & Richards, 1995). Until the most recent

version, this has involved either manual coding of whole documents for demographic information or the inclusion of keywords that could be searched as a basis for the coding of documents (which could then be 'automated' using command files as a tool for batch processing). With version 4 it became possible to import descriptive data in tabulated form ('table import'), either from a specially constructed table of data (e.g. in Excel or Word), or directly from data set up for statistical analysis (such as in SPSS). In a single step, with table import, codes for descriptive data and/or categorised responses to questions (including scales) can be created and data documents (each respondent's text responses) coded accordingly.

In a quantitatively analysed study of responses to brain death and the request to donate organs, 69 next-of-kin were asked why they had agreed to or declined to donate their relatives' organs, as well as their attitude to and rationale about organ donation for themselves. Most answers were categorized as reflecting altruism (helping another), pragmatism (not wanting to "waste good organs"; organs not healthy enough) or concern with the integrity of the body (feeling that the person had "suffered enough"; concerns about being "cut up"). Whether the respondents had agreed to organ donation or not was known. Further, the bereavement counsellor who had conducted the interviews had provided a rating of grief resolution. Those coded as pragmatic in their rationale for organ donation were generally on the road to resolution of their grief, and those who were concerned with body integrity were largely unresolved in their grief, but those who were 'altruistic' were equally split between the two ends of the grief resolution spectrum with few in the middle – about which the quantitative analysis provided no further enlightenment. Some time later, when I was introduced to NUD*IST (at that stage, version 2!), I entered all text responses and additional comments into the program, delighting, incidentally, in the capacity to enter and code text wherever it occurred regardless of whether or not it was in response to the 'right' question. The capacity to sort the text into a matrix, producing a crosstabulation between (in this case) donation values and grief resolution, and then retrieve the text for each cell within the matrix was the key to further insights. Thus it was found that, regardless of organ donation status, all of those unresolved in their grief who expressed altruism did so in absolute terms: "If other people can live, why not?", whereas all of those who were resolved in their grief expressed altruism in quality of life terms: "A man would be very selfish if he died with healthy organs and didn't give someone else a chance to lead a normal life." These were differences of potential significance that had not been previously apparent.

Techniques available in NUD*IST 4 for readily integrating descriptive coding and categorised responses with text and for displaying the results of that integration, facilitated the incorporation of additional notes made by interviewers in a study of home-based palliative care for oncology patients. The researcher involved had designed her study initially as a quantitative survey employing a (well-designed) precoded questionnaire for data collection, but 'gatekeeping' by palliative care nurses rendered the sample too small for effective statistical analysis. In an eleventh hour effort to 'dredge' whatever data were available for incorporation into a report of the study (should one ever admit to such activity?), all additional responses (identified by question numbers) were typed for each respondent. A command file was written to search the text and create nodes to store text

references for each question, allowing one-step 'automatic' coding of any responses to each question. Selected quantitative variables (demographics and responses to key questions about care and access to services) were retained from the SPSS database and imported as a table into NUD*IST, providing instant coding for each respondent on each of the imported variables. It was then possible (again, in one step) to produce a report of all text responses associated with each question, with 'coding stripes' alongside that text providing information about other responses from that person. In this case, the maximum possible 26 categories of information were employed – rather like having any of 26 colours of highlighter pen (but using letters of the alphabet) running down the side of the page next to relevant text. With just a few hours work, then, the researcher was provided with all textual information sorted (by question) and identified (by coding stripes) in such a way that she could use it not just for illustrative quoting, but rather to extend her interpretation of the responses and understanding of the issues for those with particular characteristics or experiences.

Integrating quantitative responses with qualitative data: Moving beyond simple sorting of text by descriptive or scaled categories. Importing numeric data into NUD*IST can be extended beyond the descriptive to include other precategorized and/or scaled responses, making it possible to ask quite new questions of the data. For example, numerically coded and text responses may have been initially entered together into a spreadsheet or database. Numeric codes can be transferred to SPSS for statistical analysis. Text responses can be prepared in a word processor by creating a merge file that uses the spreadsheet or database as the source of merged data. A macro is used to split the merged file into individual data documents, which are then imported into NUD*IST. Selected precategorized responses and scaled data used in the statistical analysis can also be imported into the NUD*IST database as document coding using the table import function.

In a study of discrimination among women working within a male-dominated scientific discipline, it was a simple step to sort written examples and comments from the 212 respondents about being treated differently, being harassed or experiencing discrimination according to whether the respondent gave a yes or no response to these questions. What becomes potentially more interesting is a fresh capacity to consider discrepant responses, that is, those where the example given did not reflect the yes or no categorical response selected. When is harassment (or discrimination) 'acceptable', that is, not felt to be sufficient to warrant a 'yes' response, yet warrants reporting as an example? How is this different from harassment that is clearly unacceptable? For example, "occasional sexist remarks from staff members" might be considered harassment by one respondent but not by another. Is it the nature or level of the harassment that is different, or is the assessment of harassment in these more marginal cases more to do with the respondents themselves? Do they differ from those who provide consistent responses, either demographically, or in their other categorical or text responses? Is there anything significant, for example, about the respondents who rationalized those things that others might find objectionable? One of these perceived the situation as being "unavoidable – they can't help noticing a woman in a 'man's' job." Another suggested,

> You are obviously more conspicuous than the average male. People will tend to notice you more. Some will be nicer to you. Some will patronise you. Some will

> find you easier to communicate with. Some will feel less confident of your technical ability. But whether these things amount to discrimination I don't know. I feel strongly that once people get to know you, their level of respect for you is based much more on whether you can do your job or not than on whether you are male or female.

These two respondents had differing levels of qualification, but perhaps it was significant that each was older and married (in contrast to many other respondents, and noting from the statistical analysis that those who were younger were significantly more likely to report harassment). Asking questions and analysis were both facilitated by the interdependent use of qualitative and quantitative computing.

Potentially, the matching of scores from scaled measures, for example, of anxiety, depression, or experience of pain, could be incorporated into a NUD*IST database to be assessed against interview or observational information for a subsample of respondents, in a 'dialectical' analysis (Greene & Caracelli, 1997) to ascertain the meaningfulness of those scaled scores. If combined also with the use of multiple correspondence analysis (Blasius & Thiessen, 1998), then there is indeed powerful evidence to support or refute the validity of the scale.

Quantitative analysis employing coding derived from unstructured data . As was suggested above, the capacity to fuse descriptive or scaled data with qualitative responses or interview material is not really new; instead, it has been made easier with advances in software development. But with the tools now available it has become possible to take information derived from unstructured textual (or other qualitative) data and incorporate it into a quantitative analysis, giving access to new variables and making possible new analyses.

A study of compensation paid by insurers for injuries received from motor vehicle accidents is using, as its primary data source, the judgments made by the judge or arbitrator deciding each case. These judgments include a review of the factors considered in determining the amounts awarded: as qualitative text they are quite unstructured and highly variable in content and length. While the primary analysis is qualitative, designed to derive an understanding of the process of judges' decision-making in such matters, quantitative analysis testing the relevance of particular considerations to the quantum of the total award, or components of it, also becomes possible. Thus, qualitative and quantitative data fuse to give a rounded understanding of process and outcome. If the text of the judgment includes discussion relating, say, to the veracity of the plaintiff or to the regularity of his or her previous employment or perhaps to whether the medical assessments were those of a treating doctor or one brought in especially by the insurance company or whether there is conflicting medical evidence presented, that part of the text is coded at appropriate categories in NUD*IST, regardless of where it occurs in the document. The presence or absence of such coding for each case and, indeed, whether within the unstructured text multiple codes occur in a patterned relationship or not can be exported in table format directly to a statistical database such as SPSS and added to the variables already there (which include demographics, facts about the accident, type of injury received, and so forth). Statistical testing is then being used (with a large, consecutively drawn sample) to assess whether these considerations, either singly (using t-tests) or in combination with

existing quantitative variables (using multiple regression), are associated with the level of compensation payout.

In the public health arena, it is planned to use a similar strategy as part of a review of parents' decision making about childhood immunization. Parents are being asked to describe, narrative style, the process of their thinking about immunization when their youngest child was last due for vaccination. These narratives are being qualitatively analyzed in order to develop a model of the decision-making process for this event, with particular attention being paid to the rational/less rational elements of those decisions. Coding related to specific features (the relevance of which may have been derived from prior theory, such as risk assessment, or which have become apparent from the qualitative analysis, such as needle phobia) extracted from the narratives can then be employed in quantitative analyses, such as logistic regression, with immunization compliance as the dependent variable. These analyses then provide an evaluation of the (log) odds of compliance for each feature considered by the parent, controlling for the effect of other features considered. Alternatively, it is possible to use the dichotomous codes to construct truth tables for use in Qualitative Comparative Analysis (QCA) (Ragin, 1995), a technique that produces one or more logically minimised sets of response configurations of causal variables associated with the outcome (timely immunization, in this case). QCA, which can be applied to a smaller data set than is required for multivariate statistical procedures, makes no assumptions about the nature of the variables and views causation as conjunctural rather than additive. As a consequence, a model of various alternative paths to immunization compliance may be built.

For the kinds of analyses being described here, the issue may not be so much what the judge concludes about the veracity of the plaintiff or what the parent thinks in relation to risks associated with diseases and vaccines, but rather, whether such factors even enter their respective considerations. This is quite different, then, from a traditional form of survey-derived quantitative analysis in which the respondent indicates *how* they think on a particular issue, with questions being asked about all issues deemed relevant by the researcher. The latter may be subject to all the usual biases resulting from social (professional) desirability, irrelevance to the respondent, and so on. Those particular problems inherent in survey analysis may be largely avoided by these methods while still gaining advantages that quantitative analyses can offer to add to those available through the qualitative analyses. The questions answered by such analyses are, however, different from those which might have been answered by either quantitative or qualitative analysis alone.

More complex designs currently being considered involve the application of cluster and correspondence analysis to develop models of association between concepts; the effect of weighting concepts on a qualitative analysis, and so on. Some processes use raw document by document data; others are applied to matrix data developed in and exported from the qualitative text. It is absolutely critical, in any of these processes, for the analyst to be conscious of the exact nature of the data he or she is dealing with, and to be thoughtful about the questions asked of that data and the interpretations given. Also critical, otherwise there would be no point in using a qualitative program in the task, is the importance of knowledge of the text responses themselves. These must be reviewed not only to ensure the

appropriateness of exported coding, but also to ensure a 'grounded' understanding and interpretation of the concepts and issues being considered.

Conclusion

Several clear benefits to fusing qualitative and quantitative computing in these ways are apparent:

- greater sensitivity to variations in the data (dimensionalizing of a concept) may occur as a result of 'splitting' coded material by the categories of a demographic, scaled or other quantitative response;
- new questions can be asked, in response to an apparent dialectical difference between a quantitative and a qualitative response;
- qualitative detail can be examined in the context of a broader quantitative picture;
- new sources of data can be employed to answer complex questions for which either the quantitative or the qualitative data alone would be incomplete; and
- causality (or at least, prediction) can be modelled in alternative ways.

'Good' data analysis and 'good' theory, regardless of whether the evidence is from qualitative or quantitative sources, requires a sensitivity to what the data are saying to be able to ask sensible and intuitive questions of the data in the first place. The messiness of theory development, described in the qualitative arena as a nonlinear process involving speculation, checking, selecting, revising and verifying relationships in the data (Miles & Huberman, 1994), is not restricted to qualitative social science; it is also characteristic of much laboratory science. Quantitative and qualitative procedures can be seen simply as providing alternative ways of looking at the data to explore and generate or validate theory. Indeed, both statistical analysis and laboratory science (Latour, 1987; Lynch, 1993) may require a capacity to consider the data in a manner far more in tune with a traditional view of qualitative methodology than quantitative (in so far as qualitative methodology is interpretive and inductive) but with descriptions given in numeric terms and checking of the hunches and ideas carried out using numeric rather than textual data. In this sense quantitative methods can stand *with*, rather than *beside*, qualitative data and procedures.

It is of interest to note that Lynch (1993) used the same term as Denzin & Lincoln (1994) when he wrote of "the *bricolage* of laboratory shop practice" and drew on the works of Schutz and Derrida to describe the scientist as *bricoleur*, that is, as "a jack-of-all trades who adapts 'the means at hand' – a collection of tools, scraps of material and heterogeneous skills – in trial-and-error fashion to contend with the contingencies arising in an open series of applications" (p.150). Advances in qualitative computing programs have overcome the technical barriers to combining differing data types, providing the jack-of-all-trades researcher with the tools for fresh and more insightful analyses.

References

Barbour, R. S. (1998). Mixing qualitative methods: Quality assurance or qualitative quagmire? *Qualitative Health Research, 8*(3), 352-361.

Blasius, J. & Thiessen, V. (1998). Assessing the meaning of response alternatives through multiple correspondence analysis. Paper presented at the 14th World Congress of the International Sociological Association, Montreal, July 28.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. In J. C. Greene, & V. J. Caracelli (Eds), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (pp. 19-32). San Francisco: Jossey-Bass.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, CA: Sage.

Datta, L. (1997). A pragmatic basis for mixed-method designs. In J. C. Greene, & V. J. Caracelli (Eds), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* . San Francisco: Jossey-Bass.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issues in mixed-method evaluation. In J. C. Greene, & V. J. Caracelli (Eds), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (pp. 5-18). San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. N. K. Denzin, & Y. S. Lincoln (Editors), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Kuckartz, U. (1997). winMAX: Software for Qualitative Data Analysis. Berlin: BSS

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

Lynch, M. (1993). *Scientific practice and ordinary action.* Cambridge University Press.

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

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research, 8*(3), 362-376.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*, 120-123.

Patton, M. Q. (1989). *Utilisation-focused evaluation* (2nd ed.). Newbury Park, CA: Sage.

Qualitative Solutions and Research (1997). Non-numerical Unstructured Data Indexing Searching and Theorising, Revision 4 (NUD*IST 4). Melbourne: QSR.

Ragin, C. C. (1995). Using qualitative comparative analysis to study configurations. In U. Kelle (Ed), *Computer-aided qualitative data analysis: Theory, methods and practice* (pp. 177-189). Thousand Oaks, CA: Sage.

Richards, T. & Richards, L. (1995). Using hierarchical categories in qualitative data analysis. In U. Kelle (Ed), *Computer-aided qualitative data analysis: Theory, methods and practice* (pp. 80-95). Thousand Oaks, CA: Sage.

Webb, E. J.; Campbell, D. T.; Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally