Published in final edited form as:

Field methods. 2010 May ; 22(2): 115–124. doi:10.1177/1525822X09360760.

Bridging the Qualitative/Quantitative Software Divide

Rachelle Annechino,

Prevention Research Center, Pacific Institute for Research and Evaluation, 1995 University Avenue, Suite 450, Berkeley, CA 94704, Telephone: (510) 883-5776, Fax: (510) 644-0594, rannechino@prev.org

Tamar M. J. Antin, and

Prevention Research Center, Pacific Institute for Research and Evaluation, 1995 University Avenue, Suite 450, Berkeley, CA 94704, Telephone: (510) 883-5749, Fax: (510) 644-0594, tantin@prev.org

Juliet P. Lee

Prevention Research Center, Pacific Institute for Research and Evaluation, 1995 University Avenue, Suite 450, Berkeley, CA 94704, Telephone: (510) 883-5772, Fax: (510) 644-0594, jlee@prev.org

Abstract

To compare and combine qualitative and quantitative data collected from respondents in a mixed methods study, the research team developed a relational database to merge survey responses stored and analyzed in SPSS and semistructured interview responses stored and analyzed in the qualitative software package ATLAS.ti. The process of developing the database, as well as practical considerations for researchers who may wish to use similar methods, are explored.

Keywords

quantitative methods; qualitative methods; mixed methods; methodology; databases

Introduction

Although researchers frequently use separate software packages to store and analyze qualitative and quantitative data, combining these data in one software package can also be useful for exploring and analyzing data collected in mixed method projects. In this article, we describe a relational database designed to merge survey responses stored and analyzed in SPSS with semistructured interview responses stored and analyzed in the qualitative software package ATLAS.ti (Muhr 2004). This article provides practical considerations and advantages of using this kind of database as a tool for analysis and briefly explores other uses for relational databases in research.

Data referenced in this article were collected as part of a mixed methods study of tobacco use among Southeast Asians in northern California. Field researchers conducted a confidential in-person interview with each respondent, collecting quantitative data in a brief, close-ended survey and qualitative data in an in-depth, semistructured interview. The research team was interested in comparing individual respondents' survey responses to a language-based acculturation scale to their discussions of language, culture, identity, and tobacco use in the semistructured interview component.

Background

Most mixed methods studies utilize some combination of qualitative and quantitative methods (Ames, Grube, and Moore 2000; Kinn and Curzio 2005; Happ 2006). Here, the use of the terms qualitative and quantitative refers to datasets typically generated with different primary analysis strategies in mind, even if in practice the boundaries between qualitative and quantitative datasets may be permeable (Bernard 1996; Trochim n.d.). As described by Bernard, “most quantitative analysis in the social sciences involves reducing people ... to numbers; most qualitative analysis involves reducing people to words” (2002:419). By combining qualitative and quantitative methods in one study, research questions can be informed by both approaches.

Although some software packages may be especially suited to combining selected qualitative and quantitative data, in practice the software that researchers use is rarely determined solely by fitness for a specific task. Rather, software may be selected based on factors such as availability, cost, degree of training and expertise needed, recognition in the field, and types of analyses offered by the program. Although there are programs available, such as MAXQDA, or QDA Miner with Simstat, that may perform some of the specific functions detailed here, our project opted to use ATLAS.ti and SPSS because these were the software programs that had been selected by the majority of qualitative and quantitative researchers within our research center.

The software met most of our needs admirably, but combining data directly within SPSS or ATLAS.ti posed a challenge. For example, it would be difficult using tools native to SPSS or ATLAS.ti to create a report, sorted by acculturation score and grouped by respondent, that displayed all of a respondent’s survey items related to acculturation, together with his or her discussions about acculturation. In SPSS, we could easily look at counts of codes together with other quantitative data, but without the context provided by the actual quotations. In ATLAS.ti, we could create primary document families representing our quantitative data, but because each unique value within a field would be assigned to a separate document family, we would lose the ability to easily sort and group the data as values within a single field. In addition, we would not be able to perform numeric operations on the values or display the values in the body of ATLAS.ti’s built-in query reports.

Therefore, we chose to use a Microsoft Access database as an intermediary between ATLAS.ti and SPSS. Since all three applications are widely available, one could think of the procedure outlined here as a recipe for combining data using ingredients commonly found in many ethnographic kitchens.

Using a Relational Database

Several factors influenced our decision to use a database application that supports relational database design, specifically Microsoft Access, to combine our data. Briefly, the relational database is a popular method of database design that structures data as a collection of tables that can be related to each other in complex ways, giving users a great deal of flexibility and control in managing and manipulating interrelated data (Moseley and Mead 1993; Harrington 2002; Krueger 2004).

A combination of qualitative and quantitative datasets typically involves layers of data that exist in complex relationships with each other. For example, a single respondent may have one or more sets of survey responses, together with one or more related interviews. Each interview may be broken down into multiple quotations, and each quotation may be relevant to multiple themes. Because of the nested relationships between different pieces of data, a relational database design is particularly appropriate.

In addition, incorporating coded qualitative data into a database expanded the possibilities for the use of the relational database to facilitate the exploration of relationships between qualitative and quantitative data at the level of analysis. Databases developed for integrating qualitative and quantitative data that have been described in the literature have tended to involve qualitative data consisting of relatively short text items collected via a structured protocol (Moseley and Mead 1993; Poteete and Ostrom 2008). Creating one standard database structure for unstructured qualitative interviews, however, would often be impractical. The technique described in this article works for less structured qualitative data by leveraging coding that has been done with ATLAS.ti as a way to structure the qualitative data for incorporation into the database. Using text that has been tagged with relevant codes reduces and structures the qualitative data, without reducing it to exclusively numeric representations of code frequencies, thus preserving the “qualitativeness” of individual narratives.

Combining Data in Access

For our study, the data were combined in Microsoft Access through the following process. First, the SPSS dataset was exported to an Excel file. The resulting Excel file was imported into Access with a numbered primary key field (“Respondent ID” in Figure 1), identifying each unique record in the newly imported table “SPSS_Data.” In the example shown in Figure 1, “Respondent A” is identified with the number 1 as a Respondent ID.

The next step was to import data from ATLAS.ti. An ATLAS.ti report on quotations with relevant codes—the set of qualitative data that we were interested in exploring in this case—was exported to a text file. The text file was imported into Access, with the text of quotations imported as memo fields to prevent data loss, since Access has a limit of 255 characters on text fields. On import, each row in the newly created Access table “Atlas_Import” corresponds to one line of text in the text file we created with ATLAS.ti (Figure 2 Frame A).

Using queries and the Visual Basic programming language, a procedure was written to progress through the table line by line. Each quotation was grouped into one table row, and a number corresponding to the Respondent ID shown in Figure 1 was assigned to each item. The new table created by this procedure, Atlas_Data is shown in Figure 2 Frame B. (With newer versions of ATLAS.ti, it may be possible to export the necessary data in a structured XML format, eliminating the need to group each quotation into one table row.) The quotation identified with Respondent A in Figure 2 Frame A is shown in Figure 2 Frame B with the number 1 in the Respondent ID field (“Resp_ID”), indicating that the quotation belongs to the respondent identified with the number 1 (Respondent A) in the SPSS_Data table shown in Figure 1.

A query linking the SPSS data to the ATLAS.ti data was created in the Access database (Query1 in Figure 3). Query1 instructs the database to find all the items in the SPSS_Data table, together with all the items in the Atlas_Data table with the same Respondent ID. The field RespondentID in the SPSS_Data table, and the field Resp_ID in the Atlas_Data table link the two sets of data. The relationship between the tables shown in this query mirrors the underlying relationship created between the two tables in the database. The arrow illustrates a one-to-many relationship between the tables, with the SPSS_Data table on the “one” side of the relationship, and the Atlas_Data table on the “many” side of the relationship. Thus, for each item (SPSS case) in the SPSS_Data table, there can be multiple associated items (ATLAS.ti quotations) in the Atlas_Data table. The creation of additional tables is necessary to create a fully relational design, but the relationship illustrated in Figure 3 can be thought of as the core relationship between the two datasets for this example.

A report displaying the linked data was created (Figure 4). The report is grouped on the Respondent ID used to link the two tables and shows the survey data for each respondent in the SPSS_Data table, together with all the quotations in the Atlas_Data table for each respondent.

Practical Considerations

Most database applications that support relational database design could be used to design a similar database. One of the advantages of using smaller scale database applications like Microsoft Access is that they allow for more rapid development and design modifications compared to larger-scale database applications. Relational database design makes it possible for a relatively small, easily deployable database application built with Access to handle large quantities of text from ATLAS.ti, together with large quantities of quantitative data. However, Microsoft Access cannot handle as much data, or as many simultaneous users, as larger database applications that are more complicated to deploy and to modify. The present database application has been tested with projects containing between 100 and 200 coded, semistructured interviews, together with survey data, with no negative performance issues. For those projects that still need a larger scale database, Microsoft Access can also be a good choice to begin development with, as it can be upgraded fairly easily to a larger scale database application.

Another advantage of smaller database applications like Access is that once some groundwork has been established in the form of a well-designed data structure and perhaps a few forms displaying the data of interest in a few different configurations, the interface is typically accessible to users at many different levels. Those with a basic knowledge of Microsoft Access can use its user-friendly built-in filtering and sorting options without creating queries. Results can be exported to most applications either directly or via Excel, for those who are more comfortable working with other applications, or who wish to take advantage of other programs with powerful built-in qualitative/quantitative analysis tools at various stages of exploration.

Individuals familiar with structured query language (SQL), or with graphic query building tools similar to the one that Access features for creating SQL queries can manipulate the data in any way they wish and perform quantitative and qualitative analyses within Access. Although recreating the complex statistical analyses contained within a software package like SPSS would likely be time consuming and impractical, most relational databases, including Microsoft Access, can easily support basic statistical functions. For researchers who count code occurrences, calculations based on the co-occurrence of selected quantitative data and selected codes, for example, might be of interest. An experienced Access developer with programming skills can customize a database application in virtually unlimited ways.

Discussion

Relational databases such as the one described in this article can be a useful tool in mixed methods analysis. Because mixed methods projects typically involve nested datasets, a relational database design can give researchers a great deal of flexibility in managing and manipulating interrelated data, which can, in turn, bring relationships between different kinds of data into relief.

More generally, relational databases can be useful in the management, exploration, and analysis of datasets that are related to each other, whether those datasets are qualitative, quantitative, or some combination of the two. Imagine, for example, a school-based project in which a school's students are surveyed, demographic data are collected from various

classrooms, students from selected classrooms are interviewed three times over the course of a year, and field notes on individual interviews and on classrooms are collected. In this example, the school has multiple classrooms. Each classroom has multiple students, and each classroom can have multiple field notes. Each student can have one survey, multiple interviews, and multiple field notes. Each set of field notes and each interview transcript can incorporate multiple codes/quotations of interest.

Relational databases can be structured to accommodate many kinds of complex data hierarchies, such as in the school-based scenario described above, and can be used both as a primary data collection and management tool, and as a tool for managing data from other sources, such as PDAs, and qualitative and quantitative analysis programs. Once a collection of data has been structured in this way, it can be configured, filtered, and sorted in endless ways using queries based upon the underlying relationships, and virtually any kind of cross-comparison between the datasets can be performed. Data can be incorporated into a database from numerous sources.

With any use of technology, however, there are trade-offs. Just as software designed to support investigation of qualitative data cannot produce sound qualitative analysis, database software cannot, in itself, produce good relational design, or organize data in meaningful ways. In some research contexts, the investment of time and money necessary to produce a well-designed relational database may not be possible or worthwhile. In other contexts, however, an investment in developing a relational database can save time and money in the long run and enable analyses and perspectives on data that would be impractical otherwise.

Acknowledgments

The research and preparation of this manuscript were made possible by a grant from the University of California Office of the President's Tobacco-Related Disease Research Program (TRDRP #13RT-0058, P.I. Juliet Lee). Thanks to Roland Moore for comments on this article. A special thank you to the Pacific Institute for Research and Evaluation for their support of this project.

References

- Ames GM, Grube JW, Moore RS. Social control and workplace drinking norms: A comparison of two organizational cultures. *Journal of Studies on Alcohol*. 2000; 61(2):203–219. [PubMed: 10757130]
- Bernard HR. Qualitative data, quantitative analysis. *The Cultural Anthropology Methods Journal*. 1996; 8(1):9–11.
- Bernard, HR. *Research methods in anthropology: Qualitative and quantitative approaches*. 3rd ed.. Walnut Creek, CA: AltaMira Press; 2002.
- Happ MB. Exemplars of mixed methods data combination and analysis. *Nursing Research*. 2006; 55(2-1):S43–S49. [PubMed: 16601634]
- Harrington, JL. *Relational database design clearly explained*. 2nd ed.. San Francisco: Morgan Kaufmann; 2002.
- Kinn S, Curzio J. Integrating qualitative and quantitative research methods. *Journal of Research in Nursing*. 2005; 10(3):317–336.
- Krueger B. Why you should use a relational database instead of a spreadsheet. *Cybernetics and Systems*. 2004; 35(7–8):683–696.
- Moseley LG, Mead DM. Good relations: The use of a relational database for large-scale data analysis. *Journal of Advanced Nursing*. 1993; 18(11):1795–1805. [PubMed: 8288826]
- Muhr, T. *User's Manual for ATLAS.ti 5.0*. Berlin: ATLAS.ti Scientific Software Development GmbH; 2004.
- Poteete AR, Ostrom E. Fifteen years of empirical research on collective action in natural resource management: Struggling to build large-*N* databases based on qualitative research. *World Development*. 2008; 36(1):176–195.

Trochim, WMK. [accessed March 10, 2008] The qualitative debate. Research methods knowledge base. n.d. <http://www.socialresearchmethods.net/kb/qualdeb.php>

Variable	Type	Label	Measure	Scale	Format	Values	Missing	Display Width	Decimals	Align	Font Color	Font Style	Font Size	Font Weight	Font Color	Font Style	Font Size	Font Weight
1. Dependent	Continuous	Dependent	Mean	100	Standard			10	2	Left	Black	Normal	12	Normal	Black	Normal	12	Normal
2. Dependent	Continuous	Dependent	Mean	100	Standard			10	2	Left	Black	Normal	12	Normal	Black	Normal	12	Normal
3. Dependent	Continuous	Dependent	Mean	100	Standard			10	2	Left	Black	Normal	12	Normal	Black	Normal	12	Normal
4. Dependent	Continuous	Dependent	Mean	100	Standard			10	2	Left	Black	Normal	12	Normal	Black	Normal	12	Normal
5. Dependent	Continuous	Dependent	Mean	100	Standard			10	2	Left	Black	Normal	12	Normal	Black	Normal	12	Normal

Figure 1.
SPSS Data table

The figure illustrates the process of grouping quotation lines into a single table row in the Atlas.ti software. The left window, 'Atlas_Import : Table', displays a list of text segments. The right window, 'Atlas_Data : Table', shows the same text segments grouped into a single row, with the 'Record' count changing from 9 to 2.

Resp_ID	AtlasText
1	<p>Codes: [FemaleSmk] [ReasonsSmk]</p> <p>R: There's been a big increase in female who come from Laos that are smoking.</p> <p>I: Do you have any - do you think you may have any explanation or any reason why? Do you know why at all?</p> <p>R: Y'know, I think because women are afforded with a certain kind of freedom. It's a lot more acceptable for them to smoke here. The community is different, the community's mentality is different. Where - although, y'know, I think the community tolerate female smoking here in America more so than in Laos, so even though they may try to label or put stigma to female smoking, it's widely spread, so it's become - I wouldn't say it's become a norm, but it's something - it's a behavior that is steadily gaining acceptance, or tolerance.</p>

Figure 2.
Grouping quotation lines into a single table row

Figure 3.
Query1

[illegible]

Figure 4.
Sample report page