Methods and Management of the Healthy Brain Study: A Large Multisite Qualitative Research Project

Sarah B. Laditka, PhD, 1,2 Sara J. Corwin, MPH, PhD, 3 James N. Laditka, DA, PhD, MPA, 2 Rui Liu, MPH, 4 Daniela B. Friedman, PhD, 3 Anna E. Mathews, PhD, MS, 5 and Sara Wilcox, PhD 6

Purpose of the study: To describe processes used in the Healthy Brain project to manage data collection, coding, and data distribution in a large qualitative project, conducted by researchers at 9 universities in 9 states. Design and Methods: Project management protocols included: (a) managing audiotapes and surveys to ensure data confidentiality, data tracking and distribution; (b) managing qualitative data to ensure the accuracy and confidentiality of transcription; (c) training in qualitative methods and use of qualitative software; and (d) managing participant survey data and analysis. Results: The project team coded and managed qualitative and survey data for 69 focus groups with more than 500 participants. Multiple interactive training sessions in qualitative data analysis and use of qualitative software (ATLAS.ti) were conducted. To develop a codebook, 2 teams used an open-coding process to identify codes and develop definitions; 2 team members integrated and conceptually organized these results into an initial codebook. For the audio-recordings from each research site, 2 or 3 team members hand coded 1 transcript and calculated interrater agreement (.80 or higher). Implications: Using clear protocols, participatory training sessions, team-based coding, and frequent communication among team members via e-mail and regular in-person meetings promotes effective management of large-scale qualitative research projects.

Key Words: Data collection, Multistage codebook development, Team-based coding, Focus groups, Formative research, Project management

The Healthy Brain project is establishing a science base for effective public health communication interventions to promote brain health in diverse populations (Albert et al., 2007). This project is an activity of the Healthy Aging Research Network (HAN), in collaboration with the Centers for Disease Control and Prevention (CDC). The project is a large, multisite primarily qualitative study, conducted by researchers at nine universities. The University of South Carolina is the project’s lead center, and eight other HAN sites participate (Table 1). In the initial 2 years of the study reported here (November 2005–2007), the project team conducted 69 focus groups (FGs) with more than 500 participants who were broadly representative of the U.S. population of ages 50 years and older, and also represented physicians and caregivers. This article describes processes used to manage the data from these FGs. This information may be useful to others who design and implement large qualitative studies.

Design and Methods

Core Coding Team Composition and Resources

The core coding team (hereafter team), at the Arnold School of Public Health, University of South Carolina, consisted of five faculty researchers and
two doctoral students, and was assisted by four professional staff and one graduate student. Most researchers received modest salary support, ranging from 1% to 16% of salary; one graduate student was supported at .44 full time equivalent (FTE), and two others at less than .10 FTE each. One administrative staff person was at 0.25 FTE. Most team members also participated in other aspects of the project with this support, such as developing project protocols, and arranging and conducting nine FGs at the South Carolina site. The team devoted substantial in-kind support to the project.

**Project Protocols**

The team developed protocols to manage the project: (a) management of audiotapes and participant surveys to ensure data confidentiality, tracking, and distribution to the eight other HAN sites; (b) management of qualitative data to ensure accuracy and confidentiality of transcription; (c) training in qualitative coding theory and processes, codebook development, and use of software for qualitative data management and analysis; and (d) management and analysis of participant survey data.

**FG Discussion Guide and Participant Survey**

A nine-item FG discussion guide was developed, in collaboration with the broader HAN membership and the CDC, designed to elicit participants’ attitudes about several topics related to brain health and behaviors associated with it (J. Laditka et al., 2009). FG participants also completed a 20-item survey eliciting demographic information (e.g., gender, race) and other information (e.g., dietary behaviors).

**Results**

The team interacted frequently. In-person meetings, averaging 1.5 hr, were held twice monthly in the first 2 years (November 2005–2007) and monthly in the third year (2008). Frequent e-mail communication, averaging more than 25 team-related e-mails weekly in the first 2 years and about 15 in the third year, addressed questions and helped manage the process.

**Tracking Processing of Qualitative and Survey Data**

The team developed multipart log sheets to track receipt of transcripts, participant surveys, and data distribution. A form was created in Excel to track processes, document FG characteristics, and note distinguishing features of FGs; elements of the form are shown in Table 1. One team member logged receipt of all audio files and participant surveys.

**Team Training in Qualitative Data Analysis and Use of Qualitative Software**

All sites used ATLAS.ti (version 5.0; Muhr & Friese, 2004) software to store and manage FG data. ATLAS.ti was selected as it permits on-screen coding; allows analysis of large amounts of text across multiple codes, participant responses, and group types; and facilitates electronic sharing of data files. Three in-depth interactive training sessions in qualitative data analysis and use of ATLAS.ti were conducted before beginning data analysis. During the first session (2 hr), facilitated
by two team experts in qualitative research, the team reached consensus on an approach to qualitative coding. When coding qualitative data, researchers need to make decisions about the detail of data coding. Team members agreed to code on a substantially detailed level because later analyses could readily focus on aggregates of detailed codes, whereas additional detailed coding could be added at a later stage only by recoding the many FG transcripts, which was not possible given available resources.

The second session (3 hr), with hands-on training and practice with the software, was facilitated by an experienced ATLAS.ti user. During the third session (4 hr), team members developed and reached consensus on an approach to develop the codebook conceptually and to organize it consistently with capabilities of the software. The team also discussed how ATLAS.ti would be used to organize and manage data, considering later data analysis needs.

**Transcription Processing, Coding, and Codebook Development**

Audio files were transcribed verbatim into Microsoft Word by a professional transcription service. To verify accuracy and completeness, a professional staff member compared electronic transcripts and audio recordings word by word, making corrections as needed. Transcripts were edited to remove any identifying information.

To develop the codebook, one transcript was selected at random from the African American \( (n=5) \) FGs in South Carolina and one from the White \( (n=4) \) FGs in that state. Each transcript was independently read by two coding teams, with four members each. The teams used the discussion guide as an initial framework and, when encountering a text segment that expressed a unique idea or meaning, manually marked that segment and assigned it a semantic code. Within each team, members compared their ideas and codes to determine if they arrived at similar data interpretations. During this “open coding” (Strauss & Corbin, 1990), researchers reached consensus about each code’s meaning and definition, and also about a composite code list.

Two team members integrated and conceptually organized the two code lists to form the first codebook draft. The codebook was entered into ATLAS.ti. To update the codebook, two researchers each coded the transcripts for all nine South Carolina FGs, adding new codes as they were identified.

To promote consistency and facilitate project management, four team members participated in coding. One, with extensive qualitative research expertise, was chosen to participate in determining interrater agreement for all transcripts. To determine interrater agreement, the same two team members independently coded one transcript each from every different racial/ethnic group (African American, American Indian, Asian Indian, Hispanic, non-Hispanic White, Vietnamese, Chinese) at each site. Next, these coders met face-to-face and counted, line by line, the times their codes agreed or disagreed. The number of agreements was divided by the number of disagreements plus agreements. This fraction, multiplied by 100, provided the percent agreement. Eighty percent agreement was considered evidence of consistent coding. If 80% agreement was not reached, the transcript was recoded and the percent agreement recalculated.

As the remaining FGs were coded, the codebook was updated. One team member was responsible for adding new codes in ATLAS.ti and for distributing codebook updates to the team after revisions.

In the second and third project years, the HAN sites also conducted seven FGs with physicians and seven with caregivers of persons with Alzheimer’s disease or a related disorder. To develop codebooks for these groups, the team began with the most recent version of the codebook. To determine interrater agreement, the same three team members independently coded one transcript each for the caregiver and physician FGs, and then met to review the coding and to calculate interrater agreement, which again was expected to reach .80 or greater.

Using the query tool in ATLAS.ti, all coded data were printed by code type, reviewed for accuracy, and examined for links to other codes. After all FGs were coded, the team considered whether the relatively small number of codes that were added when coding later groups might be relevant to groups that were coded earlier. The team examined this possibility and determined that the transcripts were fully coded. This result was due to the intensive level of detail in coding and also to the fact that new codes established for later transcripts represented concepts unique to the population groups in the later FGs. For example, there were unique codes for FGs of American Indians,
caregivers, and physicians; these codes did not pertain to FGs coded earlier.

As part of the coding, the team assessed the way FGs were conducted, thus contributing to the project’s quality control. In only a few instances, transcripts suggested that an FG leader may have used leading questions, possibly biasing the results. In regular conference calls among HAN sites, the principal investigator and team members emphasized the importance of minimizing this source of bias.

**Participant Survey Data Management**

The data collected from the participant survey were entered into Epi Info 6.0, a computerized data entry program. The Epi Info files were exported to the Statistical Analysis Software (SAS) version 9.1.3 (SAS Institute, 2005), in a way that allowed identification of individual FGs, HAN sites, and FGs with certain characteristics, such as race/ethnicity.

**Distribution of Qualitative and Quantitative Data**

The codebook was distributed to all HAN researchers for review and comment when it was first developed, and again following each substantive revision. Each HAN site received its coded password-protected data files by e-mail, with qualitative data in ATLAS.ti format and quantitative data in SAS format.

**Coding Team Resources**

Table 2 summarizes team resources used to manage and code the qualitative and survey data. Table 2 shows the major activities (column 2), number of team members participating (column 3), person-hours for each of the 69 FGs (column 4), and total project hours (column 5). Activities in items C–F represent fixed time costs; for these, total project person-hours indicate time costs that would be required when replicating this study approach in other projects, regardless of the number of FGs. Table 2 time estimates are conservative.

**Discussion**

Clear protocols, participatory training sessions, team-based coding, and frequent team communication by e-mail and in-person meetings promoted effective management of the Healthy Brain study. For researchers considering a large multisite qualitative research project, costs in terms of faculty, graduate student, and professional and administrative support should be carefully considered. This project yielded a large volume of social science data. However, project management, qualitative coding, and survey entry were time intensive. Time estimates presented here may guide those considering the management of a project of similar scale, to ensure that funding and institutional support are commensurate with the activity required to successfully complete the project. Additional resources to support in-person FG training with all FG moderators would enhance quality control in analogous multisite projects and would be desirable.

Several elements contributed to successful project management. First, the team made frequent face-to-face communication a priority. Face-to-face communication allowed the team to discuss and reach consensus on complex activities such as the level and specifics of coding detail. Second, in-depth training was held with all team members before data coding; these sessions helped ensure that team members participated in all coding-related decisions and promoted consensus building. Third, the team developed a comprehensive and detailed approach to ensure that each step of the process could be tracked and to develop detailed records of the project activity that were shared with all HAN sites. Fourth, the team identified one researcher to manage the codebook in ATLAS.ti and to enter all the transcript data into that software; this step promoted consistency in codebook development and maintenance. Fifth, the lead team member used frequent face-to-face, e-mail, and phone communications with all team members throughout the project to: (a) ensure management of each step, (b) identify and manage data bottlenecks that occurred when a large number of FG audio recordings or transcripts were received together, and (c) ensure that FG transcripts were distributed equitably among coders. The lead team member also communicated clearly with the eight other HAN sites about the coding process, during monthly conference calls and through e-mail. Sixth, the Healthy Brain project’s principal investigator marshaled the resources needed for the project’s success, participated as a team member, and ensured clear communication with all HAN sites. Finally, all HAN researchers, including those on the team, had substantial expertise in
qualitative research methods and analysis. Processes described in this article may help guide other research teams involved in large-scale multisite qualitative projects.

**Funding**

The research reported in this publication was supported in part by Cooperative agreements from the CDC’s Prevention Research Centers (PRC) Program HAN, Special Interest Project (SIP) 13-04 and SIP 8-06, and by Cooperative Agreements 1-U48-DP-000025, 1-U48-DP-000033, 1-U48-DP-000045, 1-U48-DP-000048, 1-U48-DP-000050, 1-U48-DP-000051, 1-U48-DP-000052, 1-U48-DP-000054, and 1-U48-DP-000059. The PRC HAN is supported by CDC’s Healthy Aging Program.

**Acknowledgments**

We thank the members of the project team for their outstanding assistance with this research: Dale Morris, Kimberly Butler, Courtney Davis, Marcia Lane, and Carol Cornman. The findings and conclusions in this study are the authors’ and do not necessarily represent the official views of the CDC or other institutions with which the authors are affiliated.

**References**


**Table 2. Estimated Person-Hours Required for the Project Management, Qualitative and Survey Coding and Data Entry for the Healthy Brain Project**

<table>
<thead>
<tr>
<th>Item</th>
<th>Activity</th>
<th>Persons participating</th>
<th>Person-hours per focus group</th>
<th>Total project person-hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Communications management with 8 other HAN sites (principal investigator)</td>
<td>1</td>
<td>9.42</td>
<td>650</td>
</tr>
<tr>
<td>B</td>
<td>Project management at the lead center</td>
<td>1</td>
<td>11.13</td>
<td>768</td>
</tr>
<tr>
<td>C</td>
<td>Development of preliminary codebook and team meeting (2 hr)</td>
<td>8</td>
<td>n.a.</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>Coding printed transcripts, 9 focus groups</td>
<td>8</td>
<td>n.a.</td>
<td>48</td>
</tr>
<tr>
<td>E</td>
<td>Basic training, ATLAS.ti (3 hr)</td>
<td>10</td>
<td>n.a.</td>
<td>30</td>
</tr>
<tr>
<td>F</td>
<td>ATLAS.ti coding strategy meeting (4 hr)</td>
<td>6</td>
<td>n.a.</td>
<td>24</td>
</tr>
<tr>
<td>G</td>
<td>Verifying transcription accuracy, data de-identification, transcript formatting</td>
<td>1</td>
<td>4.00</td>
<td>276</td>
</tr>
<tr>
<td>H</td>
<td>Entering survey data into Epi Info</td>
<td>1</td>
<td>2.50</td>
<td>173</td>
</tr>
<tr>
<td>I</td>
<td>Coding the same transcripts by 2 persons for interrater agreement, 1 transcript for each of 9 group types plus for each of 8 HAN sites</td>
<td>4</td>
<td>n.a.</td>
<td>272</td>
</tr>
<tr>
<td>J</td>
<td>Coding transcripts from 66 focus groups, coded by 1 person each</td>
<td>3</td>
<td>n.a.</td>
<td>264</td>
</tr>
<tr>
<td>K</td>
<td>Coding printed transcripts, entering into ATLAS.ti, updating codebook</td>
<td>5</td>
<td>6.00</td>
<td>414</td>
</tr>
<tr>
<td>L</td>
<td>Entering codes into ATLAS.ti and updating codebook (5 hr per focus group)</td>
<td>1</td>
<td>5.00</td>
<td>345</td>
</tr>
<tr>
<td>M</td>
<td>General team meetings, 1.5 times monthly, Years 1 and 2, monthly Year 3</td>
<td>8</td>
<td>4.87</td>
<td>336</td>
</tr>
<tr>
<td>N</td>
<td>Supplemental meetings, averaging 3 team members</td>
<td>3</td>
<td>0.87</td>
<td>60</td>
</tr>
<tr>
<td>O</td>
<td>Team e-mails to address questions and manage the process</td>
<td>10</td>
<td>7.25</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td><strong>Total project person-hours</strong></td>
<td></td>
<td></td>
<td><strong>4,176</strong></td>
</tr>
</tbody>
</table>

Notes: HAN = Healthy Aging Research Network. Item A: (5 hr total communication management per week × 50 weeks × 2 years)/69; estimate is for communications specific to focus group activities, and excludes related principal investigator effort such as project planning and evaluation, budgeting and budget management, and so forth. Item I: the nine group types included African Americans, American Indians, Asian Indians, Chinese, Hispanics, non-Hispanic Whites, Vietnamese, physicians, caregivers (includes 4 hr each for two to three coders for each focus group, plus 4-hr meeting with two to three people to review codes).

n.a. = not applicable; these categories represent fixed time costs, required regardless of the number of focus groups in any large-scale multisite project.

b Estimated 25 e-mails per each of 50 weeks Years 1 and 2, 15 e-mails per each of 50 weeks Year 3, total time for writing, reading, and responding for eight people averaging 15 min per e-mail.