METHODS REPORT

Managing Complex Multi-Case Study Evaluations: Communities Putting Prevention to Work

Heather L. Kane, Laurie W. Hinnant, Amy E. Roussel, Janice P. Tzeng, and Mary Council

December 2014

RTI Press

RTI INTERNATIONAL

About the Authors

Heather L. Kane, PhD, is a sociologist at RTI International.

Laurie W. Hinnant, PhD, is a community health psychologist at RTI International.

Amy E. Roussel, PhD, is an organizational sociologist and is the Vice President of RTI's Public Health Research Division.

Janice P. Tzeng, MPH, is a research public health analyst at RTI International.

Mary Council, BA, is a health policy analyst at RTI International.

RTI Press publication MR-0029-1412

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Kane, H. L., Hinnant, L. W., Roussel, A. E., Tzeng, J. P., & Council, M. (2014). *Managing complex multi-case study evaluations: Communities Putting Prevention to Work* (RTI Press publication No. MR-0029-1412). Research Triangle Park, NC: RTI Press. Retrieved from http://www.rti.org/rtipress

This publication is part of the RTI Methods Report series.

RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
E-mail: rtipress@rti.org
Web site: www.rti.org

DOI: http://dx.doi.org/10.3768/rtipress.2014.mr.0029.1412 **www.rti.org/rtipress**

Managing Complex Multi-Case Study Evaluations: Communities Putting Prevention to Work

Heather L. Kane, Laurie W. Hinnant, Amy E. Roussel, Janice P. Tzeng, and Mary Council

Abstract

Between 2010 and 2012, as part of the Communities Putting Prevention to Work (CPPW) initiative, the Centers for Disease Control and Prevention (CDC) funded 50 states, six US territories, and 50 communities to support high-impact, evidence-based, population-wide strategies to create healthy environments for their residents. CPPW is a locally driven initiative with a primary focus on prevention and control of tobacco use and obesity. As part of this initiative, CDC also funded an implementation evaluation to describe and understand how evidence-based, community-level improvements are applied in the field and contribute to improvements in health. We conducted this evaluation using multi-case study methods that best captured the local context and implementation processes. Large, cross-case evaluations present challenges that single or small multi-case evaluations do not. These challenges include creating flexible, but standard data collection instruments; ensuring the feasibility and utility of instruments and processes through pilot testing; promoting consistent data collection and quality; and managing a large qualitative data set and coding team. In this report, we document some strategies regarding data collection, management, and analysis that should be beneficial to other organizations supporting public health initiatives and to investigators in designing the strongest possible evaluations using large multi-case design.

Contents

RTI INTERNATIONAL

Introduction

Case studies are a widely accepted qualitative evaluation tool that enables evaluators to examine a particular case or set of cases in-depth. This approach allows evaluators to understand the unique context within which an intervention or program is being implemented and the impact that context has on the implementation process. Although common case study best practices and considerations are well documented (Patton, 2002), the methodology used to conduct case studies and the analytical process for examining and interpreting case study data vary and are evolving. According to Yin (2009), "the analysis of case study evidence is one of the least developed and most difficult aspects of doing case studies" (p. 27).

Although evaluators can draw upon these best practices, complex multi-case evaluations present unique design and analysis challenges , such as simultaneously capturing case-specific and cross-case data, coordinating multiple teams for site visits, and analyzing extremely large quantities of qualitative (and sometimes quantitative) data. In this report we highlight strategies for managing challenges in planning and implementing a large-scale case study evaluation. We present these strategies in order to help other evaluators or researchers improve processes when designing a large, complex multi-case study. These strategies are derived from our experience in conducting a large case study evaluation, the CPPW Case Study Evaluation, funded by the Centers for Disease Control and Prevention (CDC). Throughout the report, we provide examples from this evaluation to provide real-world application of these lessons.

Description of Communities Putting Prevention to Work

In 2010, the CDC funded Communities Putting Prevention to Work (CPPW) initiatives in 50 states and six US territories ("state awardees") to support high-impact, evidence-based, population-wide environmental improvement strategies. CDC also funded 50 CPPW community awardees to implement similar evidence-based, community-level improvements. The community awardees included health departments and community-based organizations working within cities, counties, tribal organizations, and larger public health jurisdictions. State and community awardees varied in how they addressed the goals of the CPPW initiative. However, all CPPW awardees sought to increase physical activity, provide and improve access to nutritious foods, decrease obesity prevalence, reduce tobacco use, and protect people from the harms associated with exposure to secondhand smoke, a known carcinogen. CPPW funding was provided through the American Recovery and Reinvestment Act of 2009 and the Affordable Care Act and gave awardees financial support unprecedented in many state and local health departments.

Through CPPW, CDC provided each state and community awardee with 2-year cooperative agreement funding for 2 years to support or augment their efforts to create healthy environments for their residents. The strategies implemented had to be drawn from the evidence base, which included the peer-reviewed literature, recommendations of the Community Preventive Services Task Force from *The Guide to Community Preventive Services* (Community Preventive Services Task Force, 2014), and subject matter expert panels and reports. Using evidence-based strategies to design a comprehensive and robust set of strategies, CPPW awardees were expected to improve health outcomes with sustainable effects within their state or locality.

Awardees implemented their CPPW strategies through partnerships with local, community, and state organizations. These include governmental agencies, private organizations and foundations, and other groups, some of which may have independent (nonfederal) resources to support advocacy and lobbying activities (e.g., American Cancer Society). Awardees advanced sustainable outcomes through education, coalition building, and partnerships. Awardees used CPPW funds to educate the public and stakeholders, disseminate information about public health problems and evidence-based solutions, and implement specific strategies (e.g., improving school meals). A comprehensive description of the CPPW initiative appears in "Fifty Communities Putting Prevention to Work: Accelerating Chronic Disease Prevention Through Policy, Systems and Environmental Change" (Bunnell et al., 2012).

Approach and Methods Used for Planning and Implementing the CPPW Case Study Evaluation

CPPW awardees' strategies involved a range of programmatic activities unfolding in different communities with unique histories and contexts. Traditional quantitative evaluation methods, such as surveying program staff or tabulating program outcomes, were not used because they would not capture *how* and *why* program staff accomplished (or failed to accomplish) program goals and objectives.

Because unpredictable events can have an impact on implementation of a program, understanding local context and its effect on implementing the strategy was central to the case study evaluation.

Figure 1 provides an overview of the methodology we used to implement this multi-case study evaluation. We first developed an analysis plan and conceptual model based on the client's key interests and the public health literature on implementing systems and environmental changes. Figure 2 shows the conceptual model used for this evaluation. We selected awardee programs for the evaluation

Figure 1. Overview of the methodological approach for the CPPW Multi-Case Study Evaluation

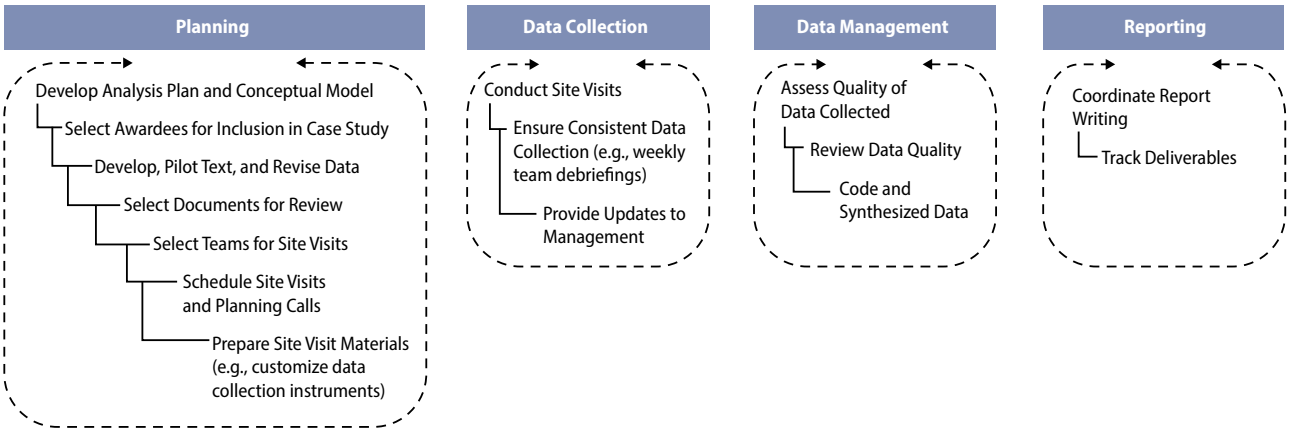
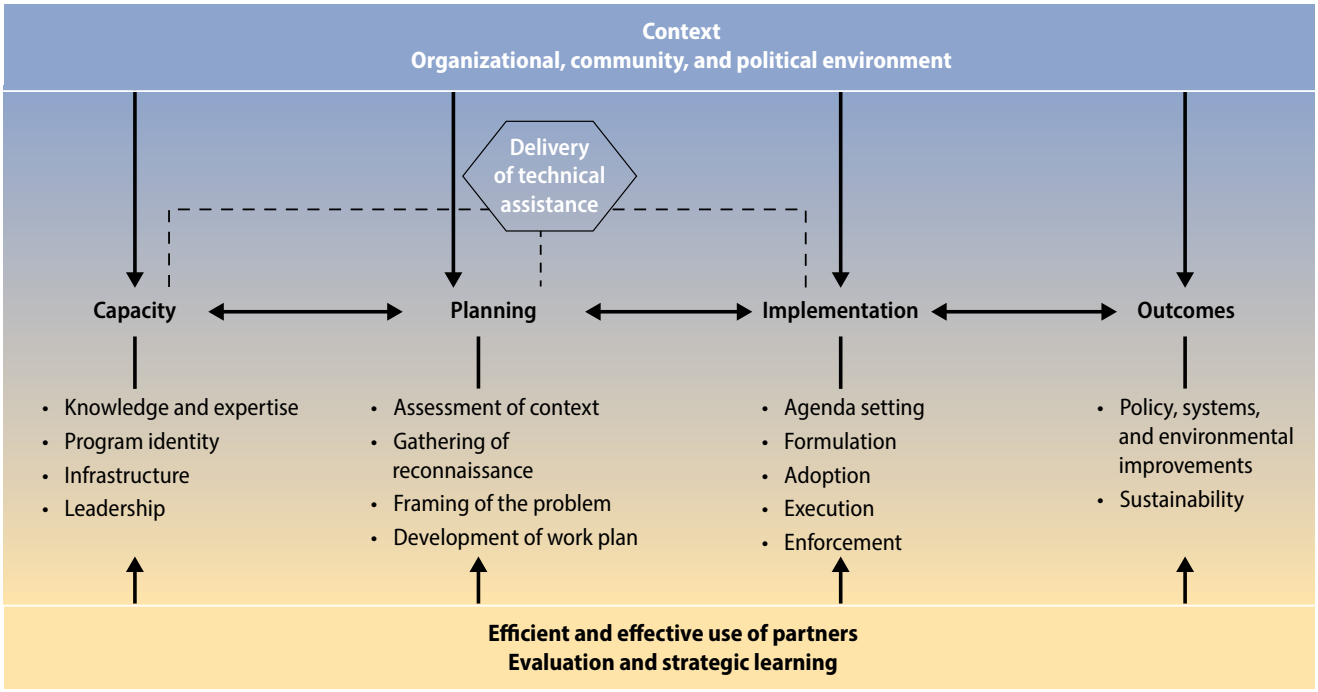


Figure 2. Conceptual framework for the Communities Putting Prevention to Work evaluation



based on several criteria: program focus (i.e., obesity, tobacco, or both), CDC award type (i.e., small city/rural, large city, urban area, or tribe), geographic region (using US Census regions), experience (calculated using a community's history of involvement with other related CDC-funded initiatives), priority populations (e.g., African American, American Indian, Hispanic), and site alignment with CPPW emerging strategic priority areas.

Using our conceptual model, we developed interview protocols for different awardee program stakeholders, including protocols for the principal investigator/program director, program staff, program partners, leadership team/advisory council members, legislators, and evaluators. Prior to the first round of site visits, two site visit teams pilot tested our instruments and site visit processes and provided feedback on revising them. Upon making those refinements, we trained site visit teams on the instruments and processes, and the project manager, with the help of each site visit team, scheduled planning calls and site visit dates with the awardees.

For the CPPW evaluation, we conducted two rounds of site visits with 22 awardee programs—the first between November 2010 and July 2011 and the second between December 2011 and June 2012. The site visits averaged 3 days in length. Each site visit team consisted of a designated interviewer and a note taker. At each site, the site visit teams conducted interviews with up to 20 program staff and leadership, key partners, and stakeholders. All interviews were audiotaped (with the key informants' permission); in addition, note takers entered interview data directly into a word processing program. Note takers were instructed to take notes as close to verbatim as possible. They used the audio recordings to complete any missing information from the interview notes. In total, 828 interviews were conducted at awardee programs.

For the cross-case analysis that looked across all 22 cases, we developed an extensive codebook derived from our conceptual model and experience during site visits. Twelve coders, most of whom were also part of the data collection team (i.e., an interviewer or note taker), were trained on the codebook and discussed codes on an ongoing basis through weekly meetings and a listserv. Coders were assigned interviews for an entire case, and where possible, a coder was often assigned to a case where he or she had served as a part of the site visit team, although this was not always possible. They coded all interview data in NVivo 9.0, a qualitative research software package produced by QSR International. A data manager assessed intercoder reliability using the kappa coefficient, and a level of 0.75 of agreement was maintained on all codes. Coding and intercoder agreement are described on pages 7 and 8 under Managing a Large Qualitative Data Set and Analysis Team.

The data manager prepared NVivo coding reports, which the coders and project leadership reviewed and synthesized for emerging themes. Themes evolved out of the data when respondents from several awardees expressed a similar idea (such as capacity for implementing changes, resources required to implement programs). Coders clearly documented the emerging themes into tables for each content area and noted the number of awardees for which the emerging theme appeared (i.e., whether it was a frequent, or strong, theme vs. a less frequent, or weaker, one). The analysis team then compiled the tables and used them in preparing a cross-case report for CDC.

Strategies for Managing Challenges of Large Multi-Case Evaluations

Large multi-case study evaluations present several data collection, management, and analysis challenges that do not typically occur in single or small case study evaluations. In this section, we describe some of the challenges and how we managed the challenge in the CPPW case study evaluation. In particular, we detail how we (1) collected standard data elements while also customizing data collection to the diverse cases, (2) ensured the feasibility and utility of data collection instruments and processes, (3) promoted consistent data collection and quality across a large data collection team, and (4) managed a large qualitative data set and analysis team.

Collecting Standard Data Elements While Customizing Data Collection Instruments

Typically in evaluation, evaluators apply a standard instrument across all cases. This allows evaluators to collect common information across different cases so that comparisons can be made and common themes identified. Yet large-scale case study evaluations often entail examining heterogeneous cases. Even when the evaluation involves studying the implementation of similar programs (e.g., a standard curriculum), the context of each program, the roles of staff, and strategies for implementation can vary widely. The heterogeneity that occurs in real-life situations necessitates flexible data collection instruments that allow evaluators to capture unique contexts and dynamic processes. At the same time, though, those instruments must ensure collection of comparable information.

Managing the challenge of simultaneously gathering case-specific, but comparable information involves relying on the evaluation questions to establish core topics for exploration and adapting instruments to align with the case context. To accomplish this balance, evaluators can use the evaluation conceptual framework, evaluation planning matrix, and evaluation questions to identify important topics (e.g., capacity for implementing changes, resources required to implement programs) to cover when examining each case and then develop a generic data collection instrument or set of instruments based on the key informant's role. For example, separate instruments may be used for the principal investigator, program staff, and partner organizations. Before collecting data (e.g., conducting a key informant interviews), evaluators should review program documents or other available information (e.g., media coverage) to modify the generic questions to align with the case and the roles of individuals involved.

In the CPPW case study evaluation, awardee programs implemented diverse strategies such as educating the public on the dangers of secondhand smoke, collaborating with schools to improve school meals, and working with county leaders to develop master plans that promote physical activity. A single instrument or survey could not capture the diversity of the implementation strategies, key partners, or contextual factors, nor could a single instrument gather adequate information on *how* awardee program staff implemented public health improvements.

To obtain information across this extremely diverse group, we developed a semistructured qualitative interview guide that was standardized in terms of addressing the key evaluation questions and topics of interest but was tailored to the selected communities, strategies, and key informants included in the case study. We crafted the generic interview guide based upon two primary guidance documents: (1) the case study evaluation questions and (2) the CPPW case study conceptual framework (see Figure 2). By ensuring the questions aligned with the evaluation questions and conceptual framework, we focused on information needed to meet the priorities of the evaluation, and the need to ask questions that address the unique strategies being implemented, the community context, and the role of the respondent.

For example, the CPPW generic interview guide for program staff queried: "Prior to obtaining CPPW funding, what was your community/state doing to address policy, systems and environmental improvements related to tobacco and/or obesity?" To tailor the question for one awardee program, we reviewed the program application, which included background on the awardee program's experience. The staff in that awardee program had worked on getting bike lanes included in three towns' transportation plans prior to receiving CPPW funding. We revised the original question to ask how that experience supported their CPPW efforts and prompted for any other efforts not listed in the application.

Although this approach balances gathering comparable and case-specific data, it requires advanced preparation and staff time before the site visit. The amount of labor time depends on several factors, such as the amount of textual materials available before the site visit and the complexity of the case and data collection instrument. In spite of this disadvantage, this approach does offer another advantage for data collection. Tailoring each interview guide ensures that each interview is an

efficient use of the informant's and interviewer's time by eliminating questions that are not relevant to the informant.

Ensuring the Feasibility and Utility of Data Collection Instruments and Processes

Just as survey developers extensively test instruments prior to implementation, pilot testing the data collection instruments and site visit processes (e.g., applying scheduling strategies, identifying roles of key informants) allows an evaluation team to identify and address potential pitfalls before full implementation. In case study evaluations, the evaluation team usually has access to some information about the case prior to data collection. This information often includes program applications, evaluation plans, and logic models. However, these materials tend to be outdated by the time they reach the evaluation team because changing conditions or on-the-ground realities require that an awardee make changes in work plans, staffing, or partners.

For the CPPW evaluation we conducted a pilot site visit with two awardees 3 months before full implementation began. This allowed for sufficient time to make modifications to the process and materials and obtain CDC approval for the changes. We conducted extensive debriefing after each day of pilot interviews to review the data collected, discuss interview questions that worked well or those that needed to be reframed, consider which staff to interview given the questions of interests, and weigh the number of questions relative to the length of time available for each interview.

We used the information to adapt the proposed processes, guides, and training for the project team. For example, prior to our first site visit, we planned to interview people in particular roles (e.g., program staff, program director, principal investigator, partners, local evaluator). While on site, we learned that financial administrative staff were critical to the implementation of this time-sensitive initiative. In planning the remaining site visits, we included financial administrative staff on our site visit schedules. Although not every project has the flexibility to incorporate time for testing processes and materials through pilot site visits, they can be beneficial in helping project management to identify challenges early on, test interview guide questions and improve the team's understanding of what to expect on each visit.

Promoting Consistent Data Collection and Quality Across a Large Data Collection Team

Multi-case evaluations often require large data collection teams, especially if the cases are geographically dispersed and if a tight timeline drives data collection. With large data collection teams, comprehensive training on data collection and management can improve data quality. Individuals comprising a large data collection team often come with different experiences and expectations about how to collect, record, and store data. Some individuals think bullet point notes on interviews are adequate; others strive for near-verbatim interview notes. Even when data collection instruments seem obvious, individuals have unique perspectives on the purpose of the questions.

Taking the time to formally train the data collection team prior to initiating data collection will helps to ensure that all members are familiar with the data collection and management procedures to be used, the priorities of the evaluation, anticipated challenges, and strategies to overcome those challenges. Data collection trainings can often be done in a group setting and use virtual communication resources, such as a webinar, video conference or teleconference for staff who may not be located in one site. Where appropriate, one-on-one trainings can be helpful.

For the CPPW evaluation, project leaders conducted mock interviews with site visit leads to ensure lead interviewers were familiar with the data collection instruments they would be using and how to probe appropriately on challenging topic areas. We held a specific training for note takers to ensure notes were recorded consistently (i.e., noting the question asked and response, and taking note as close to verbatim as possible rather than paraphrasing) to minimize discrepancies in the note taking process. This was especially important for CPPW because the notes served as the primary data file for cross-case analysis (as opposed to a transcript from a recording).

In addition to the individual trainings, we developed a large group training that outlined the initiative,

provided an overview of the site visit planning and logistics, and described data collection processes. To help prepare team members for the training, we gave them pre-training materials (i.e., FAQs about the project, a glossary of teams, sample interview guides) to review. Having team members look over these materials before the training allowed for more interaction during the training sessions. During the actual training, we shared specific guidance on the data collection instruments and their purpose, how to tailor data collection instruments, and how to name and where to save data files (e.g., audio files, interview notes) and described the intended timing of completing pre- and post-site visit activities (e.g., when to tailor instruments, when to complete the site visit report).

Managing a Large Qualitative Data Set and Analysis Team

The layers of complexity for data management and analysis for large multi-case evaluations necessitate careful data management and a sizeable data coding team to code the thousands of pages of interview notes or transcripts. Recognizing these needs and addressing them early in the planning process, before data collection begins, will help ease the burden on those responsible for managing the large amount of data that will be coming in and those who will be part of the data analysis team.

Good data management entails establishing how data will be saved while the site visit team is on site and after the site visit and developing quality assurance procedures to ensure that all teams follow data management guidelines. For CPPW, prior to the site visits, we carefully documented file naming conventions, timelines, and file server project folders for all data. All site visit teams received the documentation during training, and two designated individuals on the project team reviewed each site visit team's files after the site visit to ensure that the site visit team had completed and properly saved all files.

Analysis of large qualitative data sets also requires several key considerations. First, such large evaluations can produce thousands of pages of textual data that a handful of people could not analyze within typical project timelines (e.g., 4 to 5 months for analysis and report writing). Using a large coding team necessitates careful monitoring of individual progress and analytic files. Second, maintaining quality analysis, as measured by intercoder reliability, with a large coding team is much more challenging than with a smaller coding team. Coding qualitative data depends on applying "an abbreviation or symbol . . . to a segment of words—most often a sentence or paragraph of transcribed field notes—in order to classify the words" (Miles & Huberman, 1994, p. 56). Based on the coder's understanding of the code definition and the text he or she reviews, the coder must determine when to apply a code. Because individual judgment is involved in the application of a code, coders can disagree or have differing understandings of a code definition or when to apply the code. The more individuals on an analysis team, the greater the likelihood for disagreement or differing understandings.

For CPPW, we completed 828 interviews and relied on a coding team of 12 individuals to complete coding efficiently. With many interviews and many coders, careful data management proved critical. To ensure careful monitoring and management, we had a single, dedicated data manager. This individual assigned coders a subset of the interview files to code, merged the coded data back into the master file, and followed up on coding progress to ensure that we met timelines. Having one individual responsible for overseeing the master qualitative data file had an additional benefit. She maintained sole control of the master qualitative data file and monitored its integrity. Had multiple coders worked in the same file, it could have been easily corrupted or modified. (Corruption of a qualitative data file could result in the loss of many staff labor hours as the file cannot always be restored.) The data manager also managed backup of the master file in the event that something did happen.

To ensure that we maintained analytic quality with a large coding team, we used carefully trained coders and assessed intercoder reliability. The initial training included an overview of the qualitative research software to be used, a review of the timeline for coding and milestones for coding completion, and an extensive review of the codebook to ensure that each individual had a common understanding of the

codebook and how the codes were to be applied. We discussed areas where there was the need for double coding (applying two or more codes to the same piece of data), how each code related to the evaluation questions of interest and the client's priorities, and how much data should be coded to ensure that there was sufficient context to help explain the response.

In addition to the initial training, we maintained ongoing communication among coding team members to address questions related to coding, to review coding practices and ground rules, and to share insights or ideas generated during the coding process. We held weekly meetings to discuss the codebook and emerging challenges and to adjudicate difficult coding decisions. We also used a listserv, which allowed coders to pose real-time questions to other coders.

To assess coding consistency across multiple coders, 20 percent of the interviews were double coded; the data manager calculated a kappa statistic to determine intercoder reliability.[1] The data manager and coders ensured that the double-coded texts maintained an average kappa statistic of at least 0.75, a generally acceptable cutoff for intercoder reliability (Lombard, Snyder-Duch, & Bracken, 2002). For interviews with an average kappa statistic of less than 0.75, the two coders, the data manager, and a project leader met to reconcile the coding differences until the kappa statistic surpassed 0.75.

[1] We chose to use the kappa coefficient over percent agreement because percent agreement tends to overestimate agreement; however, we acknowledge that the kappa statistic, which uses expected agreement in its calculation, can be unreliable, especially for rare events, and has an underlying assumption of coder independence (Viera & Garrett, 2005). Despite its limitations, it is commonly used in behavioral research (Bakeman, 2000).

Discussion

Through the CPPW initiative, awardees sought to implement evidence-based public health strategies. To understand the process of implementation, an evaluation had to be designed to capture the unique contexts and approaches used by diverse awardees. Traditional evaluation strategies, including observational studies or surveys, were insufficient to capture the variety of data and situations encountered as part of the CPPW initiative and address evaluation questions of interest. A large, multi-case study methodology represented a scientifically rigorous and defensible evaluation approach. This report documents our strategies to effectively manage an extensive multi-case study, which resulted in a cross-case analysis of data from 828 interviews across 24 state and community awardees.

Large, multi-case evaluations present challenges that single or small multi-case evaluations do not. These challenges include creating flexible, but standard data collection instruments; ensuring feasibility and utility of instruments and processes through pilot testing; promoting consistent data collection and quality; and managing a large qualitative data set and coding team. We hope that our strategies regarding data collection, management, and analysis are beneficial to other organizations supporting public health initiatives and to investigators in designing the strongest possible evaluations using large multi-case design.

References

Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judge (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138–159). New York, NY: Cambridge University Press.

Bunnell R., O'Neil D., Soler R., Payne, R., Giles, W. H., Collins, J., Bauer, U., Communities Putting Prevention to Work Program Group. (2012). Fifty communities putting prevention to work: Accelerating chronic disease prevention through policy, systems and environmental change. *Journal of Community Health. 37*(5), 1081–1090.

Community Preventive Services Task Force. (2014). The guide to community preventive services. Washington, DC: U.S. Department of Health and Human Services. Retrieved October 27, 2014, from www.thecommunityguide.org

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.

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

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Acknowledgments

The authors would like to thank Kristine Day, MPH, and Robin E. Soler, PhD, both of the Centers for Disease Control and Prevention Division of Community Health, for their ongoing support and contributions to this work and manuscript.

www.rti.org/rtipress **RTI Press publication MR-0029-1412**