Research Article

A Framework for Rigorous Qualitative Research as a Component of Mixed Method Rapid-Cycle Evaluation

Qualitative Health Research
2019, Vol. 29(2) 279–289

DOI: 10.1177/1049732318795675
journals.sagepub.com/home/qhr
SAGE

Megan Skillman[1], Caitlin Cross-Barnet[2], Rachel Friedman Singer[1], Christina Rotondo[1], Sarah Ruiz[3], and Adil Moiduddin[1]

Abstract

As federal, state, and local governments continue to test innovative approaches to health care delivery, the ability to produce timely and reliable evidence of what works and why it works is crucial. There is limited literature on methodological approaches to rapid-cycle qualitative research. The purpose of this article is to describe the advantages and limitations of a broadly applicable framework for in-depth qualitative analysis placed within a larger rapid-cycle, multisite, mixed-method evaluation. This evaluation included multiple cycles of primary qualitative data collection and quarterly and annual reporting. Several strategies allowed us to be adaptable while remaining rigorous; these included planning for multiple waves of qualitative coding, a hybrid inductive/deductive approach informed by a cross-program evaluation framework, and use of a large team with specific program expertise. Lessons from this evaluation can inform researchers and evaluators functioning in rapid assessment or rapid-cycle evaluation contexts.

Keywords

research evaluation; research design; health care; mixed methods; qualitative methods; evaluation; qualitative; North America

Background

Ongoing policy changes have spurred innovations in health care delivery, increasing the need for quick turnaround in understanding program impacts, implementation successes, sustainability, and replicability. Rapid-cycle evaluation provides timely feedback to funding organizations and to program staff and care providers; offers support for continuous quality improvement (McNall & Foster-Fishman, 2007; Shrank, 2013); and allows observations of changes over time, overcoming a typical dilemma of one-time observation in traditional case study research (Schneeweiss, Shrank, Ruhl, & Maclure, 2015; Willis, 2010). Tension between time-consuming and resource-intensive grounded theory analysis (Charmaz & Bryant, 2016) and demands of rapid-cycle evaluation mean that researchers must balance rigor with tight timelines (Beebe, 2008; McNall & Foster-Fishman, 2007).

Based on our experiences evaluating the Center for Medicare and Medicaid Innovation's (CMMI) Health Care Innovation Awards (HCIA) initiative, we describe how a combined inductive and deductive qualitative approach enabled cross-program analysis as part of a rapid-cycle, multisite, mixed-method evaluation. We describe lessons learned around data collection, preparation, and analysis that are particularly relevant to how we maintained a rigorous analytic approach across multiple waves of data collection. This required managing a large volume of qualitative data and a large team so that we could generate interim and final findings for CMMI and HCIA awardees.

Researchers are adopting a range of rapid-cycle evaluation and assessment approaches entailing quantitative analyses (Cody & Asher, 2014; Shrank, 2013), qualitative analyses (Beebe, 1995; McMullen et al., 2011; Needle et al., 2003; Rowa-Dewar et al., 2008), and mixed-method designs (Bate & Robert, 2002; Chin et al., 2004; McNall & Foster-Fishman, 2007; Munoz-Plaza et al., 2016). An exclusive focus on impacts can lead to perceptions that interventions lose effectiveness when scaled up (Parry, Carson-Stevens, Luff, McPherson, & Goldmann, 2013). Conversely, purely qualitative approaches typically focus

[1]NORC at the University of Chicago, Chicago, IL, USA
[2]Center for Medicare & Medicaid Innovation, Baltimore, MD, USA
[3]U.S. Department of Health and Human Services, Washington, DC, USA

Corresponding Author:
Megan Skillman, NORC at the University of Chicago, 55 E. Monroe St., 30th Fl. Chicago, IL 60603, USA.
Email: Skillman-Megan@norc.org

on arriving at a preliminary understanding of issues or are limited to analyzing a narrow set of themes (Ash et al., 2008; Beebe, 2008). Because rapid-cycle evaluation is designed to generate insights about program implementation as well as impact, combined quantitative and qualitative analyses that map implementation and contextual factors to quantifiable impacts are optimal.

Prior work in rapid-cycle and rapid assessment qualitative research tends to apply to a narrow set of conditions. Analytic tools such as the Plan-Do-Study-Act (PDSA) cycle (Deming, 1986), rigorous and accelerated data reduction (RADaR) (Watkins, 2017), and the rapid appraisal framework (Beebe, 1995) have key limitations. Some methods lose efficiency when applied to a large volume of data (Watkins, 2017); thus, such studies tend to pertain to a specific program, issue, or condition (Brown, Lhussier, Dalkin, & Eaton, 2018; Koch & Iliffe, 2010; Murray, 1999; Rowa-Dewar et al., 2008; Schneeweiss et al., 2015; Taylor et al., 2014; Trotter, Needle, Goosby, Bates, & Singer, 2001). Other recently developed resources such as Moore et al.'s (2015) Medical Research Council guidance on process evaluation offer practical application of theory to real-world research; however, there is little information on how to work within a rapid-cycle context. We offer insight into the broad, practical applicability of our methods across a diverse portfolio of programs and add to the growing body of literature based on recent interest in rapid analyses (Neal, Neal, VanDyke, & Kornbluh, 2015; Vindrola-Padros & Vindrola-Padros, 2018; Watkins, 2017).

The HCIA

In 2012, CMMI awarded US$900 million to 107 programs across the country to test new payment and service delivery models with the aim to lower costs and improve health and quality of care for Medicare, Medicaid/Children's Health Insurance Program (CHIP), and dually eligible enrollees (Centers for Medicare & Medicaid Services, 2016). As illustrated in Table 1, the 18 HCIA programs in our evaluation, implemented from 2013 to 2016, focused on seven conditions issued CMMI priority status based on their associated costs, prevalence, seriousness, and the potential impact of improved treatment. Each awardee's program had goals to improve clinical processes, intermediate clinical outcomes, and patient quality of life as well as to reduce acute health care utilization and costs for the target condition. Programs also had multidisciplinary care teams to manage patients across several settings including primary care, outpatient care, inpatient hospital care, home health care, emergency rooms, community health centers, community-based mental health centers, skilled nursing facilities, and rehabilitation hospitals. Descriptions of interventions studied and the results of our evaluation have been previously published (NORC at the University of Chicago, 2016, 2017).

Evaluation Framework

Our goals in this mixed method evaluation were to (a) assess whether each of the awardees improved quality of care and health outcomes and lowered health care costs and (b) identify how program elements such as organizational structure, geographic location, and workforce characteristics influence awardee outcomes. The evaluation design was based on a conceptual framework developed by Berry et al. (2013) for the HCIA initiative; the framework included implementation of innovations, outcomes (better health and quality of care, lower cost), workforce transformation (satisfaction, supply, quality), and sustainability of HCIA programs. According to this framework, the health status and characteristics of target populations, workforce (e.g., training), internal context (e.g., organizational features), and external context (e.g., overarching legal, regulatory, and fiscal environment) influence implementation effectiveness, which, in turn, influences patient outcomes. Implementation effectiveness had three dimensions that included program drivers (theory of change and theory of action underlying the program), the intervention (components, dosage, fidelity, and self-monitoring), and program reach (coverage and timeliness of implementation) (Berry et al., 2013).

Ultimately, our major research domains and subdomains included innovation program components and characteristics, targets (individual and organizational level), implementation process (self-monitoring), implementation effectiveness (fidelity, reach), context (endogenous and exogenous factors), workforce development (deployment, use of community health workers), and patient outcomes (quality of care, utilization/cost, quality of life). We followed this evaluation framework in addition to a more detailed list of domains and subdomains based on Berry et al. (2013) that was provided by the meta-evaluator for all 107 HCIA programs (Day et al., 2015). We also developed domain subthemes (e.g., provider engagement under the workforce domain) and used the resulting analytic evaluation framework to inform interview protocols, reports, and qualitative coding schemes.

The framework and approach used in this evaluation are largely in line with Moore et al. (2015) who describe how mixed methods might be appropriate to address process evaluation research questions. In particular, quantitative analyses can measure key process variables and test hypotheses whereas qualitative methods "capture emerging changes in implementation, experiences of the interventions and unanticipated or complex causal pathways, or generate new hypotheses" (Moore et al., 2015,

Table 1. Summary of HCIA Awardees and Interventions.

Disease	Awardee	Intervention	Awardee Type	Frontline Staff
Cancer	Innovative Oncology Business Solutions, Inc.	Care coordination; symptom management	Entity created for HCIA partnering with seven private community oncology practices	Triage nurses (RNs, LPNs)
	University of Alabama at Birmingham	Navigation for all cancer stages	Not-for-profit hospital system with partners across five states	LHWs, RNs, physicians, administrators
	Trustees of the University of Pennsylvania	Care coordination; symptom management; in-home support	University-based health system	NP, nurses, LHWs, social workers, chaplain
	The Rector and Visitors of the University of Virginia	Symptom management for advanced cancer	Private university health system	RN, palliative care physicians, social worker, radiation oncologist, medical assistant
Cardiovascular disease	Christiana Care Health Services, Inc.	Post-acute care; long-term care coordination	Not-for-profit private hospital system	Care coordinators (RNs), social workers, LHW, pharmacist
	Upper San Juan Health Service District	Community outreach; telemedicine; paramedicine; care coordination	Rural medical center	ER doctors, LHWs, paramedics, neurologists
Chronic pain	Mountain Area Health Education Center, Inc.	Case management; care coordination	Not-for-profit integrated health center with four sites	NPs, LHWs, pharmacists, behavioral health providers, physicians
Dementia	Trustees of Indiana University	Individualized and integrated care management	Academic/university medical center	LHWs, nurses (RNs), social workers
	Regents of the University of California, Los Angeles	Comprehensive dementia care	Not-for-profit medical clinic part of university system	Care coordinators (NPs), LHWs
Diabetes	FirstVitals Health and Wellness, Inc.	Chronic disease management	For-profit health and wellness company	Nurses (RNs), LHWs, social worker
	Joslin Diabetes Center, Inc.	Community-based diabetes education	Diabetes research and clinical care organization	LHWs, nurses, registered dietitians
	Southeastern Diabetes Initiative	Chronic disease management	Federally qualified health centers	RNs, CHWs, LPNs, CNA, PA, RN, social worker, endocrinologist, counselor, dietitians
End-stage renal disease	George Washington University	Remote telemonitoring	Academic institution	Nurses
Pediatric asthma	Health Resources in Action, Inc.	Chronic disease management	Not-for-profit community health and medical foundation with seven sites	CHWs, RNs, NP, registered respiratory therapist, physicians
	Le Bonheur Community Health and Well-Being	Chronic disease management	Not-for-profit hospital	RNs, CHWs, NP, physician, respiratory therapists, social workers
	Nemours Foundation	Chronic disease management	Not-for-profit integrated health system	CHWs, behavioral health providers, nurse
Stroke	Ochsner Clinic Foundation	Home-based follow-up care; targeted stroke education	Not-for-profit acute-care hospital	LHWs, RNs, APN
Various conditions	Vanderbilt University Medical Center	Clinical care coordination	Not-for-profit academic/ university medical center	RNs, social worker, medical assistants

Note. HCIA = Health Care Innovation Awards; RN = registered nurse; LPN = licensed practical nurse; LHW = lay health worker; NP = nurse practitioner; ER = emergency room; CHW = community health worker; CNA = certified nursing assistant; PA = physician assistant; APN = advanced practice nurse.

Table 2. Summary of Qualitative Data Collection.

Qualitative Research Mode	Timing and Sample	Purpose
Document review	Review of awardee applications, operational plans, quarterly reports (including implementation metrics, narrative progress reports, work plans, monitoring measures)	Gain baseline understanding of program scope, purpose, and ongoing progress
Baseline key informant phone interviews	Semistructured phone interviews with leaders from each awardee organization Sample: Leaders across all 18 programs	Gain baseline understanding of program scope and purpose
Site visit key informant interviews	10–15 interviews per program conducted during site visits Sample: Participants of key stakeholder groups involved in the intervention including leaders, frontline staff, site managers, providers, informatics teams, monitoring teams, workforce training teams, and partner organizations	Gain understanding of program implementation processes, challenges, facilitators, and outcomes
Patient/caregiver focus groups and phone interviews	One or two patient or patient/caregiver focus groups conducted during site visits to each program; when focus groups could not occur (generally because of participants' mobility or accessibility restrictions), individual patient/caregiver interviews were conducted Sample: Approximately 4–9 participants reflecting important clinical characteristics, demographics, and payer mix of the awardee population	Gain understanding of program impacts and implementation from patient and caregiver perspectives

p. 4), an apt description of our ultimate approach. To determine the unique purposes of qualitative and quantitative analyses in the evaluation, we considered (a) what types of qualitative and quantitative data we anticipated having available to us, (b) our research questions, and (c) the evaluation framework. Our approach also had elements of an outcome evaluation in that quantitative analyses of Medicare and Medicaid claims and encounter data assessed impacts on cost, quality, and utilization. In addition, our qualitative team sought to collect data about and understand patients' experiences with programs including patients' self-reported improvements in health, quality of life and care, and utilization.

Data Sources and Methods

We conducted two rounds of site-based data collection from each program that resulted in more than 300 interviews with staff, leaders, and partners (559 individuals); 40 patient or patient caregiver focus groups (311 individuals); and 134 one-on-one patient or caregiver interviews. All caregiver/patient focus groups and most staff interviews occurred in person, but some interviews occurred via phone to accommodate schedules or health needs. Interview and focus group participants provided verbal consent to participate and be recorded. We conducted all interviews and focus groups between March 2014 and December 2015. Table 2 provides more detail on qualitative data collection; NORC at the University of Chicago's Institutional Review Board approved this study.

In addition, we corresponded with programs approximately monthly to discuss emerging challenges, clarify understanding of programmatic functions, exchange information about claims or survey data quality, and gather updates on implementation status following in-person site visits or telephone interviews. Formal rapid-cycle feedback occurred on a quarterly and annual basis through written reports to CMMI that CMMI then shared with HCIA awardees.

Findings

A number of factors and decisions enabled us to conduct rapid and rigorous qualitative analysis with a large volume of data. From the beginning of the evaluation, our team anticipated the volume of qualitative data to be collected from the 18 HCIA programs. We focused qualitative analyses on answering specific research questions around implementation and impact, starting with basic program information and implementation experiences. In line with a grounded theory approach, data collection and analysis occurred simultaneously. We designed workflows so that our analysis informed subsequent data collection efforts (Charmaz & Bryant, 2016; Merriam & Tisdell, 2015); however, we also deductively shaped data collection instruments and a codebook based on Berry et al.'s (2013) conceptual framework. Table 3 illustrates facilitators and lessons learned from this experience.

Adapting to Rapid-Cycle Analysis Through Planning

The ongoing feedback loop intrinsic to rapid-cycle evaluation required efficiency. Over the course of two years, we

Table 3. Summary of Facilitators and Lessons Learned.

Evaluation Stage	Facilitators and Lessons Learned
Planning and data collection	• Preparation of detailed notes instead of transcripts • Staging multiple waves of coding
Analysis	• Hybrid inductive/deductive approach • Cross-program research questions • Simultaneous coder training and codebook refinement • Coder expertise on specific programs • A large team trained to implement quality assurance procedures

generated qualitative analyses for six quarterly and three annual reports. Specific CMMI interests and the availability of quantitative and qualitative data informed the focus of quarterly reports; we used annual reports to present summative information across programs. Figure 1 shows the interaction between primary and secondary data collection, qualitative coding and analysis, quantitative analysis, and evaluation reporting. As recommended by Miles, Huberman, and Saldana (2013), we applied lessons learned from data collection to inform analysis and vice versa.

Given the number of HCIA programs in our portfolio and the volume of data collected, we formed a 15-person qualitative research team. Two doctoral-level qualitative researchers with substantial experience (a health policy expert trained in qualitative research and a cultural anthropologist) led the team and determined the overall qualitative methodological approach. Five individuals with graduate degrees in policy and several years of health research experience managed data collection among HCIA awardees. These individuals focused on subgroups of awardees (e.g., all cancer-focused awardees, all diabetes-focused awardees); their engagement with a specific subset of programs over multiple years was necessary to have a rich, continually evolving understanding of HCIA interventions (Morse, 2015) and enhanced our ability to draw comparisons across relatively similar programs. A coding manager with significant academic health policy research experience led the qualitative coding effort. Seven junior researchers, all with at least a bachelor's degree, assisted with both data collection and coding.

Data preparation remained a flexible process to allow us to test various approaches. We employed our own rapid-cycle improvement approach to data preparation based on ongoing lessons learned. Using professional transcription companies during the first round of data collection (March–October 2014) added burden on the research team, who had to redo much of the work because the transcriptionists had difficulty tracking multiple speakers and correctly capturing medical terminology. Therefore, in the second round of site visits (February–June 2015), trained note-takers on the evaluation team created detailed notes that conveyed all thoughts expressed in each interview, then compared notes to interview recordings as a quality assurance measure. This made content clearer and easier to code while eliminating the need for professionally created transcripts.

Staging qualitative coding allowed the team to keep pace with ongoing primary data collection and quantitative analyses. A rapid-cycle structure cannot accommodate delaying development of codes and an analytic framework until after data collection has been completed. We found that coding data in multiple, smaller rounds gave the team flexibility to conduct ongoing qualitative analysis. To complement quantitative analyses presented in annual and quarterly reports, we scheduled three separate waves of coding. During two waves, each timed after a round of in-person site visits, we focused on implementation effectiveness. The third and final wave of coding focused on program effectiveness. For all waves, we used the same group of seven trained coders.

As part of the first wave of coding, we examined the experiences of frontline staff to corroborate assertions made by program leadership and validate our baseline understanding of programs. In Wave 2, we built on of our understanding of what staff were doing to analyze how effectively different program components (e.g., care coordination, home visits) were operationalized within interventions. Timing of the third wave of coding corresponded with delivery of quantitative data as awardees closed out their programs. In this wave, we coded all patient and caregiver focus groups from both rounds of site visits, which contained detailed information about patient outcomes and program effectiveness.

Because this third wave took place later in the evaluation, we had a better understanding of outcome-related themes and quantitative results. As a result, we could refine analyses across both qualitative and quantitative teams to explain agreement or differences between the qualitative and quantitative data. Rather than revisiting data to capture entirely new themes, we added codes and subcodes (e.g., divided one patient outcomes code into subcodes about perceptions of utilization and changes in attitudes and behaviors). Researchers who design multiple stages of coding and analysis can efficiently use time and resources but should note that periodic codebook revisions can be problematic if they are too dramatic and require recoding data. Shifting focus throughout this evaluation and reshaping hypotheses still required time and energy. In an effort to learn as the evaluation progressed, we made small but important refinements to the codebook and repeated coder training processes before each wave of coding to maintain interrater reliability (IRR).

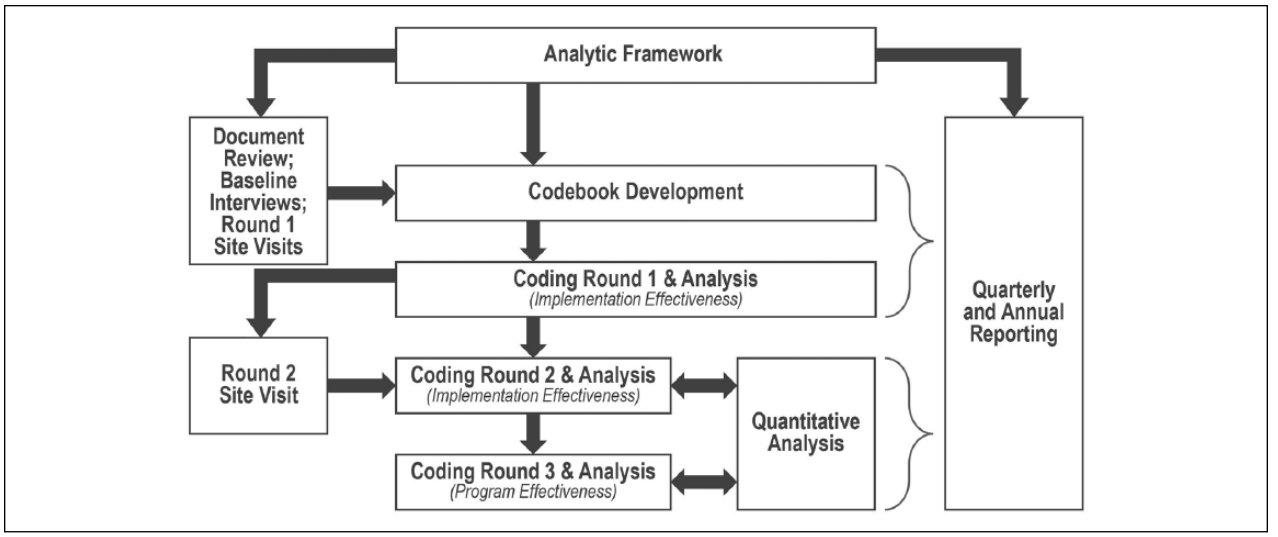

Figure 1. HCIA evaluation data collection, analysis, and reporting.
Note. HCIA = Health Care Innovation Awards.

An Analytic Approach for a Rapid-Cycle Environment

Rapid assessment or rapid appraisal methods typically entail inductive analysis (Beebe, 1995; Rowa-Dewar et al., 2008), an exploratory approach (Stebbins, 2008) that develops hypotheses based on existing facts (Benaquisto, 2008). Because we adhered to a conceptual framework and prescriptive research questions and needed to report findings before all data collection cycles were finished, a wholly inductive analysis was not possible. Conversely, employing a purely deductive analysis with qualitative data was not ideal because it can lead to bias or prevent new themes from emerging (Hsieh & Shannon, 2005; Tracy, 2012).

Hybrid inductive/deductive approaches have been noted for their adaptability in various settings or to diverse materials (Fereday & Muir-Cochrane, 2006; Garrison, Cleveland-Innes, Koole, & Kappelman, 2006; Morgan, 2007). Combining approaches was useful given the diversity of the 18 HCIA programs in our portfolio and the need to generate findings both at the individual program level and across programs early on in the evaluation. We used deductive analysis to identify topics of interest, hypothesize, observe, and then confirm (Yardley & Marks, 2004) based on the evaluation's conceptual framework, then applied principles of inductive analysis to refine and expand our analytic themes. After we identified an initial set of codes derived from the evaluation framework research domains, codebook development followed a more inductive, iterative process in line with best practices (Beebe, 2008; Glaser & Strauss, 1967). Filling in codebook "gaps" with inductively identified codes enabled us to detect emergent themes from interview data (Bradley, Curry, & Devers, 2007; Corbin & Strauss, 1990; Garrison et al., 2006).

Applying an agile framework helped standardize analyses across diverse programs. The conceptual framework research domains in this evaluation were flexible enough to enable analysis (a) within a single program and (b) across a group of similar programs or program components (e.g., use of home visits or health information technology) even if program aims, context, target populations, or disease foci differed. As discussed by Jenkins, Slemon, Haines-Saah, and Oliffe (2018), purposefully keeping analytic domains (and codes) broad generated a degree of data reduction and enabled program-specific analysis within each domain. We aligned semistructured interview protocols and the codebook with the evaluation's conceptual framework domains to collect and classify data consistently across all programs. In addition, we created interview questions and codes under the domains concerning short- (e.g., program enrollment), intermediate- (e.g., multidisciplinary teamwork), and long-term impacts (e.g., improvements in patients' health) to provide benchmarks throughout the life of the project and offer rapid-cycle feedback to awardees.

During interviews, HCIA intervention staff reconsidered their goals and targets throughout the program duration, prompting them to adjust as needed. For instance, after holding a patient focus group, we found that patients were not familiar with a low-touch program or its functions. As a result, program staff sought to clarify their role with patients. Often, demographic interview questions about patient or staff characteristics required interviewees

to reflect on who they were serving and whether that group was the appropriate target population.

We designed interview protocols to apply to all programs but tailored them to fit each program's unique components, workface, context, and theory of change. Given our knowledge of how each research domain was standardized or tailored with respect to interview protocols, we could anticipate which codes would be challenging to apply across programs and sought to emphasize them during coder training. Our final codebook included approximately 30 codes spread across the following topics: implementation effectiveness (e.g., timeliness of intervention delivery), workforce roles and training, health and utilization outcomes, policy and market context, awardee organizational characteristics, and challenges and facilitators. Codes were not mutually exclusive, meaning that coders could apply multiple codes to the same section of interview text as necessary.

Simultaneous codebook refinement and coder training was more efficient than staging each separately. Teams can save time by simultaneously refining a codebook and training coders (Skillman, Singer, Rotondo, Ruiz, & Moiduddin, 2017). Because of the diversity of interventions in our portfolio, the codebook had to be tested across a range of programs and a large team facilitated such testing. We began training coders after qualitative team leads agreed on major research domains and a set of preliminary codes. From this point, coders provided direct input on the codebook. Although simultaneous codebook testing and coder training generated many questions, required garnering coder buy-in, and occasionally resulted in confusion because of the volume of code changes, coders gained a sophisticated understanding of concepts because they had followed the evolution of codes over time. These steps required an engaged coding team committed to tracking updates and would not work well with a less engaged group. In addition, centralizing communications about code changes and questions on a single coding manager created a system for consistent messaging across the team.

When coders identified data that did not fit into existing codes, the coding manager reviewed the data and noted trends. When necessary, we created new codes, collapsed codes, or refined code definitions. In this way, triangulating perspectives of a large team strengthened and added validity to these decisions.

Although, theoretically, any individual should be able to use a codebook independent of background knowledge, some interviewees referenced information that required a deep understanding of the program's context to ascertain the speaker's main point. For instance, some interviewees described lessons learned from interventions implemented before HCIA funding, which, for non-experts, appeared indistinguishable from relevant program data. We made sure that all coders practiced coding data from programs for which their level of familiarity varied. When necessary, program experts annotated transcripts to clarify details for training purposes. Thus, coders saw how the same codes applied in multiple contexts (e.g., a "dosage" code meant home visits in one program and telephone calls in another).

Forming a large team that harnessed coders' background knowledge of specific programs conserved the time and resources it would have required for coders to become familiar with all programs. We trained coders within a 2- to 3-month period. Each coder had to achieve desired IRR (70%) with at least two other coders before independently coding transcripts and detailed notes. Whenever possible, coders cleaned and coded transcripts (or their own detailed notes) from site visits they had attended. Because of the large volume of data we had on each intervention, leveraging coders' specialized knowledge about particular programs saved substantial coding time. Beebe (2001) cautions that teams of more than six individuals are less likely to be productive than smaller teams during group interactions. We protected against this by conducting highly structured team meetings and sharing new coding decisions in writing. Coders individually coded then met in rotating pairs to compare coding schemes over multiple rounds of practice. This was critical to generating a consistent coding style across the team. We calculated IRR with Cohen's Kappa on independently coded transcripts and detailed notes using *NVivo* software (version 10; QSR International, Doncaster, Victoria, Australia). Although there is no universally accepted Kappa value (Bakeman, McArthur, Quera, & Robinson, 1997), we set an IRR benchmark at 70%, which falls in the middle of the range of Kappa scores considered "substantial agreement" (Bakeman, 2000; Viera & Garret, 2005). We deemed this threshold appropriate given the team's varying programmatic expertise and the complexity of interview and focus group data.

Six months after the second round of site visits, coders replicated major elements of the training outlined above until the group reached 70% or higher IRR. Throughout this process, coders were conducting additional site visits necessitated by evaluation's rapid-cycle structure. Assuming an active role in data collection enhanced coders' ability to understand how interview content evolved over time and, consequently, enhanced their ability to propose meaningful refinements to the codebook. The challenging but effective steps we followed, meaning (a) arranging logistics for concurrent data collection and analysis, (b) determining appropriate IRR thresholds, and (c) conducting highly structured training, might apply to other multisite, rapid-cycle evaluations. Maintaining coding quality over an extended period of time was also essential for validity and needed particular attention, which we discuss below.

Ongoing quality assurance exercises helped coders maintain IRR. Throughout all waves of independent coding, team members resolved questions by conferring with one another and participating in additional rounds of IRR calculations to ensure that reliability remained high. The coding manager also randomly spot-checked coding in each wave, escalated coding questions to program experts, and tracked coding questions to verify that decisions were consistent. The larger evaluation team added another layer of quality assurance to this process by reviewing findings derived from coded data and clarifying nuances.

Discussion

This article fits into a growing body of work focused on the use of qualitative methods to develop actionable findings that can be used to support implementation and policy decisions (Munoz-Plaza et al., 2016; Schneeweiss et al., 2015). Compared with academic social science, applied fields including health services and policy research have always had greater need for combined deductive and inductive approaches as well as more timely qualitative and mixed method results. In addition, the introduction of new payment and service delivery models has increased the demand for rapid-cycle qualitative analysis to understand how organizations are implementing policies and programs, the unintended consequences of implementation, and external factors influencing implementation.

Methods described previously have generally focused on evaluating a specific program or topic (Koch & Iliffe, 2010; Murray, 1999; Rowa-Dewar et al., 2008; Schneeweiss et al., 2015; Taylor et al., 2014; Trotter et al., 2001). Our analysis entailed evaluation of a diverse set of programs in an effort to understand each one and identify general lessons about innovations in health care. A wholly inductive approach to qualitative data analysis involves multiple iterations of data collection and analysis including three levels of coding (Corbin & Strauss, 1990) that is more suitable for a small team of researchers with a limited focus, a long timeline, and researcher-driven objectives. Even though our work drew from principles of grounded theory, the overall approach and analysis was driven by a pragmatic perspective (e.g., Morgan, 2007), a preformed evaluation framework, and a set of prescriptive research questions. Recognizing that research is rarely either purely inductive or deductive, Morgan (2007) discusses a pragmatic paradigmatic approach that relies on a combination of induction and deduction, which he calls "abduction." Researchers adapting or applying our methods should engage in reflexivity by considering the underlying assumptions of the research as well as associated implications for methods and potential bias (Morgan, 2007; Morse, 2015).

There are several limitations to using the methods presented here. First, this rapid-cycle strategy required a large and relatively stable team of researchers. Optimally, coders participate in primary data collection, data cleaning, coding, and analysis, which means they must be able to travel and must possess a broad skill set. Our qualitative team consisted of 15 people, 13 of whom remained involved for the length of the project. Such involvement is likely not possible for projects with limited funding. Because some turnover is inevitable, quality assurance measures must be established and implemented to maintain rigor over time. In addition, our methods meant sacrificing a purely inductive approach; some researchers voice concern that using deductive analysis with qualitative data can lead to bias and stifle creativity (Hsieh & Shannon, 2005; Tracy, 2012). Finally, despite the rapid-cycle nature of the methods we present here, qualitative data collection and analysis is time consuming, especially compared with the time needed to analyze an established quantitative data set such as Medicare claims. Our methods overcame challenges typical of qualitative research regarding timing, specifically in that some analysis was conducted before all data collection was complete.

Although this article focuses on evaluation planning, qualitative data collection, and analysis, we recognize broader considerations for conducting a mixed method rapid-cycle evaluation. For instance, more exploration is needed into the unique methodological challenges of how different quantitative and qualitative data analysis timeframes and units of analysis impact how findings are reported. Rapid-cycle feedback might lead intervention staff to make midstream changes or might lead policy makers to intervene based on early results before fully understanding the intervention and the mechanisms behind its impacts, both of which can impact the integrity of the evaluation. Given these considerations, a rigorous, flexible approach to program evaluation that leverages the strategies we have described will provide a foundation for researchers to deliver timely and comprehensive findings that can inform program implementation and policy change.

Authors' Note

This work was completed while Dr. Ruiz was a Senior Research Scientist at NORC at the University of Chicago.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

This work was supported by the Center for Medicare & Medicaid Innovation (Contract no. HHSM-500-2011-00002I). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services or any of its agencies.

References

Ash, J. S., Sittig, D. F., McMullen, C. K., Guappone, K., Dykstra, R., & Carpenter, J. (2008). A rapid assessment process for clinical informatics interventions. In *American Medical Informatics Association Annual Symposium Proceedings* (pp. 26–30). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656056/pdf/amia-0026-s2008.pdf

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

Bakeman, R., McArthur, D., Quera, V., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods*, *2*, 357–370. doi:10.1037/1082-989X.2.4.357

Bate, P., & Robert, G. (2002). Studying health care "quality" qualitatively: The dilemmas and tensions between different forms of evaluation research within the UK National Health Service. *Qualitative Health Research*, *12*, 966–981. doi:10.1177/10-4973202129120386

Beebe, J. (1995). Basic concepts and techniques of rapid appraisal. *Human Organization*, *54*, 42–51. doi:10.17730/humo.54.1.k84tv883mr2756l3

Beebe, J. (2001). *Rapid assessment process: An introduction*. Lanham: AltaMira Press.

Beebe, J. (2008). Rapid assessment process. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 726–728). Thousand Oaks, CA: Sage.

Benaquisto, L. (2008). Coding frame. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 86–89). Thousand Oaks, CA: Sage.

Berry, S. H., Concannon, T. W., Gonzalez-Morganti, K., Auerbach, D. I., Beckett, M. K., Chen, P. G., . . . Weinick, R. M. (2013). *CMS Innovation Center Health Care Innovation Awards evaluation plan*. RAND. Retrieved from https://www.rand.org/pubs/research_reports/RR376.html

Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*, *42*, 1758–1772. doi:10.1111/j.1475-6773.2006.00684.x

Brown, S., Lhussier, M., Dalkin, S. M., & Eaton, S. (2018). Care planning: What works, for whom, and in what circumstances? A rapid realist review. *Qualitative Health Research*, *28*, 2250–2266.

Centers for Medicare & Medicaid Services. (2016). *Health Care Innovation Awards*. Retrieved from https://innovation.cms.gov/initiatives/Health-Care-Innovation-Awards/

Charmaz, K., & Bryant, A. (2016). Constructing grounded theory analyses. In D. Silverman (Ed.), *Qualitative research* (pp. 347–360). Thousand Oaks, CA: Sage.

Chin, M. H., Cook, S., Drum, M. L., Jin, L., Guillen, M., Humikowski, C. A., . . . Schaefer, C. T. (2004). Improving diabetes care in Midwest community health centers with the health disparities collaborative. *Diabetes Care*, *27*, 2–8. doi:10.2337/diacare.27.1.2

Cody, S., & Asher, A. (2014). *Smarter, better, faster: The potential for predictive analytics and rapid-cycle evaluation to improve program development and outcomes*. Retrieved from https://www.brookings.edu/research/smarter-better-faster-the-potential-for-predictive-analytics-and-rapid-cycle-evaluation-to-improve-program-development-and-outcomes/

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*, 3–21. doi:10.1007/BF00988593

Day, J., Bir, A., Kawati, L., Kane, H., Smith, K., Stockdale, H., . . . Ayub, A. (2015). *Health Care Innovation Awards meta-analysis and evaluators collaborative annual report year 1*. RTI. Retrieved from https://innovation.cms.gov/Files/reports/hcia-metaanalysis-evalcollab.pdf

Deming, W. E. (1986). *Out of the crisis*. Cambridge: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*, 80–92. doi:10.1177/160940690600500107

Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *The Internet and Higher Education*, *9*, 1–8. doi:10.1016/j.iheduc.2005.11.001

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Aldine Transaction.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*, 1277–1288. doi:10.1177/1049732305276687

Jenkins, E. K., Slemon, A., Haines-Saah, R. J., & Oliffe, J. (2018). A guide to multisite qualitative analysis. *Qualitative Health Research*, *28*, 1969–1977. doi:10.1177/1049732318786703

Koch, T., & Iliffe, S. (2010). Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: A systematic review. *BMC Family Practice*, *11*, Article 52. doi:10.1186/1471-2296-11-52

McMullen, C. K., Ash, J. S., Sittig, D. F., Bunce, A., Guappone, K., Dykstra, R., . . . Wright, A. (2011). Rapid assessment of clinical information systems in the healthcare setting. *Methods of Information in Medicine*, *50*, 299–307. doi:10.3414/ME10-01-0042

McNall, M., & Foster-Fishman, P. G. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation*, *28*, 151–168. doi:10.1177/1098214007300895

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco: John Wiley.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis* (3rd ed.). Thousand Oaks, CA: Sage.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., . . . Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal*, *350*, Article h1258. doi:10.1136/bmj.h1258

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*, 48–76. doi:10.1177/2345678906292462

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, *25*, 1212–1222. doi:10.1177/1049732-315588501

Munoz-Plaza, C. E., Parry, C., Hahn, E. E., Tang, T., Nguyen, H. Q., Gould, M. K., . . . Sharp, A. L. (2016). Integrating qualitative research methods into care improvement efforts within a learning health system: Addressing antibiotic overuse. *Health Research Policy and Systems*, *14*, Article 63. doi:10.1186/s12961-016-0122-3

Murray, S. A. (1999). Experiences with "rapid appraisal" in primary care: Involving the public in assessing health needs, orientating staff, and educating medical students. *British Medical Journal*, *318*, 440–444. doi:10.1136/bmj.318.7181.440

Neal, J. W., Neal, Z. P., VanDyke, E., & Kornbluh, M. (2015). Expediting the analysis of qualitative data in evaluation: A procedure for the rapid identification of themes from audio recordings (RITA). *American Journal of Evaluation*, *36*, 118–132. doi:10.1177/1098214014536601

Needle, R. H., Trotter, R. T., Singer, M., Bates, C., Page, J. B., Metzger, D., & Marcelin, L. H. (2003). Rapid assessment of the HIV/AIDS crisis in racial and ethnic minority communities: An approach for timely community interventions. *American Journal of Public Health*, *93*, 970–979.

NORC at the University of Chicago. (2016). *Third annual report: HCIA disease-specific evaluation*. Retrieved from https://downloads.cms.gov/files/cmmi/hcia-diseasespecific-thirdannualrpt.pdf

NORC at the University of Chicago. (2017). *Third annual report addendum: HCIA disease-specific evaluation*. Retrieved from https://innovation.cm-s.gov/Files/reports/hcia-diseasespecific-thirdannrpt-addendum.pdf

Parry, G. J., Carson-Stevens, A., Luff, D. F., McPherson, M. E., & Goldmann, D. A. (2013). Recommendations for evaluation of health care improvement initiatives. *Academic Pediatrics*, *13*, S23–S30. doi:10.1016/j.acap.2013.04.007

Rowa-Dewar, N., Ager, W., Ryan, K., Hargan, I., Hubbard, G., & Kearney, N. (2008). Using a rapid appraisal approach in a nationwide, multisite public involvement study in Scotland. *Qualitative Health Research*, *18*, 863–869. doi:10.1177/1049732308318735

Schneeweiss, S., Shrank, W. H., Ruhl, M., & Maclure, M. (2015). Decision-making aligned with rapid-cycle evaluation in health care. *International Journal of Technology Assessment in Health Care*, *31*, 214–222. doi:10.1017/S0266462315000410

Shrank, W. (2013). The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Affairs*, *32*, 807–812. doi:10.1377/hlthaff.2013.0216

Skillman, M., Singer, R., Rotondo, C., Ruiz, S., & Moiduddin, A. (2017). *Is rigorous qualitative research possible in a rapid-cycle evaluation context?* [White paper]. Retrieved from http://www.norc.org/PDFs/Working%20Paper%20Series/WP-2017-005.pdf

Stebbins, R. A. (2008). Exploratory research. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 1327–1329). Thousand Oaks, CA: Sage.

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Quality & Safety*, *23*, 290–298. doi:10.1136/bmjqs-2013-001862

Tracy, S. J. (2012). The toxic and mythical combination of a deductive writing logic for inductive qualitative research. *Departures in Critical Qualitative Research*, *1*, 109–141. doi:10.1525/qcr.2012.1.1.109

Trotter, R. T., Needle, R. H., Goosby, E., Bates, C., & Singer, M. (2001). A methodological model for rapid assessment, response, and evaluation: The RARE program in public health. *Field Methods*, *13*, 137–159. doi:10.1177/1525822X0101300202

Viera, A., & Garret, J. M. (2005). Understanding interobserver agreement: The Kappa Statistic. *Family Medicine*, *35*, 360–363.

Vindrola-Padros, C., & Vindrola-Padros, B. (2018). Quick and dirty? A systematic review of the use of rapid ethnographies in healthcare organisation and delivery. *BMJ Quality & Safety*, *27*, 321-330. doi:10.1136/bmjqs-2017-007226

Watkins, D. C. (2017). Rapid and rigorous qualitative data analysis: The "RADaR" technique for applied research. *International Journal of Qualitative Methods*, *16*, 1–9. doi:10.1177/1609406917712131

Willis, E. M. (2010). The problem of time in ethnographic health care research. *Qualitative Health Research*, *20*, 556–564. doi:10.1177/1049732310361243

Yardley, L., & Marks, D. F. (2004). Introduction to research methods in clinical and health psychology. In L. Yardley & D. F. Marks (Eds.), *Research methods for clinical and health psychology* (pp. 1–20). Thousand Oaks, CA: Sage.

Author Biographies

Megan Skillman, M.P.A., is a research scientist at NORC at the University of Chicago, where she leads and contributes to the design and execution of qualitative and mixed method activities. Skillman has extensive experience in qualitative methods including cross-case analysis and has led qualitative analyses for several federal and foundation funded projects.

Caitlin Cross-Barnet, Ph.D., is a social science research analyst at the Centers for Medicare & Medicaid Services Innovation Centers and an associate at the Hopkins Population Center at Johns Hopkins University. Dr. Cross-Barnet's research focuses on maternal-child health, families, substance abuse and social inequality.

Rachel Friedman Singer, Ph.D., M.P.H., M.P.A., is a Senior Research Scientist at NORC at the University of Chicago. Dr. Singer has led qualitative data collection and analysis efforts for numerous large-scale, multi-site national evaluations. Prior to joining NORC, Dr. Singer focused on quality improvement in home health care while working at the Center for Home Care Policy and Research at the Visiting Nurse Service of New York.

Christina Rotondo, M.P.P., is a Senior Research Analyst at NORC at the University of Chicago, where she contributes to qualitative research activities as a part of multiple large-scale national evaluations. Rotondo has experience managing and implementing qualitative coding processes to inform mixed-methods evaluations.

Sarah Ruiz, Ph.D., is a Program Officer at the Administration for Community Living in the National Institute on Disability, Independent Living, and Rehabilitation Research. Dr. Ruiz has more than 15 years of experience designing and analyzing research studies on health outcomes for older adults and people with disabilities, with extensive experience working on projects at the federal, state and local level. Her current research focuses on evidence-based programs and practices for individuals aging with long-term disability.

Adil Moiduddin, M.P.P, is a Vice President of the Health Care Department at NORC at the University of Chicago, with over 18 years of experience in health services research, consulting and program evaluation. Moiduddin has extensive knowledge of federal and state programs to support health care financing and delivery with particular expertise in the health care safety net and health IT programs.