Published in final edited form as:

Med Care Res Rev. 2020 October ; 77(5): 416–427. doi:10.1177/1077558718803859.

A Rigorous Approach to Large-Scale Elicitation and Analysis of Patient Narratives

Mark Schlesinger¹, Rachel Grob², Dale Shaller³, Steven C. Martino⁴, Andrew M. Parker⁴, Lise Rybowski⁵, Melissa L. Finucane⁴, Jennifer L. Cerully⁴

¹Yale University, New Haven, CT, USA

²University of Wisconsin Law School, Madison, WI, USA

³Shaller Consulting Group, Stillwater, MN, USA

⁴RAND, Pittsburgh, PA, USA

⁵The Severyn Group, Ashburn, VA, USA

Abstract

Patient narratives have emerged as promising vehicles for making health care more responsive by helping clinicians to better understand their patients' expectations, perceptions, or concerns and encouraging consumers to engage with information about quality. A growing number of websites incorporate patients' comments. But existing comments have fragmentary content, fail to represent less vocal patients, and can be manipulated to "manage" providers' reputations. In this article, we offer the first empirical test of the proposition that patient narratives can be elicited rigorously and reliably using a five-question protocol that can be incorporated into large-scale patient experience surveys. We tested whether elicited narratives about outpatient care are complete (report all facets of patient experience), balanced (convey an accurate mix of positive and negative events), meaningful (have a coherent storyline), and representative (draw fulsome narratives from all relevant subsets of patients). The tested protocol is strong on balance and representativeness, more mixed on completeness and meaningfulness.

Keywords

medical consumerism; online ratings; patient narratives; star ratings; physician report cards

Article reuse guidelines: sagepub.com/journals-permissions

Corresponding Author: Mark Schlesinger, Room 304 LEPH, Yale School of Public Health, 60 College Street, New Haven, CT 06520, USA. mark.schlesinger@yale.edu.

Our findings were presented in several sessions of the AHRQ Research Conference held in Arlington, VA, in October 2015, and at an invitational stakeholder meeting on advancing the science of patient narratives sponsored by AHRQ in Rockville, MD, in September 2016.

Declaration of Conflicting Interests

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

Supplemental Material

Supplemental material for this article is available online.

Introduction

Patient narratives have recently emerged as promising vehicles for making health care more responsive to patient experiences. Their potential has been identified by health services researchers around the world (Greaves, Millett, & Nuki, 2014; Hao, 2015; Kemp et al., 2016), drawing particular attention within the United States (Dohan, Garrett, Rendle, Halley, & Abramson, 2016; Schlesinger, Shaller, & Grob, 2015). Narrative accounts identify aspects of care not captured in conventional patient experience surveys (Bardach et al., 2015; Huppertz & Smith, 2014; Ranard et al., 2016), help clinicians better understand patients' expectations and concerns (Dohan et al., 2016; Friedberg, Steelfisher, Karp, & Schneider, 2011; Riiskjaer, Ammentorp, & Kofoed, 2012; Trigg, 2011; Tsianakas et al., 2012), and encourage consumers to engage with information about health care quality (Kanouse, Schlesinger, Shaller, Martino, & Rybowski, 2016; Riiskjaer et al., 2012).

Yet while Medicare, most private insurers, and many regional health improvement collaboratives collect information on patient experience, they generally give short shrift to narrative accounts. When open-ended questions are incorporated into existing patient experience surveys, they involve largely perfunctory requests for "additional comments" (Schlesinger, Grob, et al., 2015). The resulting responses are typically omitted from public reports and, until recently, were rarely included in feedback reports distributed to clinicians (Cognetta-Rieke & Guney, 2014).

The consequences of this neglect involve more than lost opportunities. Filling a vacuum, commercial websites such as Yelp and Angie's List, which present patients' accounts of interactions with clinicians, have proliferated over the past decade (Lagu et al., 2017; Ranard et al., 2016; Schlesinger, Grob, et al., 2015). These volunteered comments are purely anecdotal, representing neither the variety of patients treated by each provider nor the full range of experiences for each patient (Cognetta-Rieke & Guney, 2014). Moreover, providers sometimes attempt to manipulate website content to burnish their reputations, further undermining the value and credibility of the posted information (Sun, 2015). If consumers assess and select providers primarily based on these published anecdotes, they may be reacting to accounts that are unrepresentative or distorted in their presentation.

Despite these liabilities, substantial marketing budgets and easier accessibility make it likely that commercial websites will increasingly "crowd-out" other forms of quality reporting (Sick & Abraham, 2011). Already more Americans consult anecdote-based websites than they do sites that report clinical quality metrics when selecting a clinician (Tompson et al., 2014). Although many are guarded in their assessment of the patient comments that they find, a substantial number of consumers pay serious attention to this information (Holliday, Kachalia, Meyer, & Sequist, 2017).

One way to protect against these problematic consequences and to take greater advantage of narratives' positive potential would be to integrate the rigorous collection of open-ended accounts into existing large-scale surveys that ask millions of Americans annually about their experiences as patients (Schlesinger, Grob, et al., 2015). For narratives to be more than accumulated anecdotes, they must be elicited from the same kind of representative sample of

patients that is carefully constructed for extant patient experience surveys (Dohan et al., 2016). Effective integration into existing surveys, however, requires designing the wording and sequencing of open-ended questions used to elicit patient narratives (henceforth, the “elicitation protocol”) and fielding them with the same scientific rigor previously applied to closed-ended questions.

Elsewhere, we have identified and operationalized criteria for assessing the quality of elicited patient narratives and described rigorous methods for determining how elicitation protocols perform with regard to these criteria (Grob et al., 2016). In this article, we assess the performance of an elicitation protocol tested with a nationally representative group of patients reporting on experiences with ambulatory clinical care. More specifically, we tested the capacity of the protocol to elicit narratives that cover the full scope of patient experience, provide a balanced assessment of the positive and negative aspects of those experiences, and convey a storyline that is coherent to readers. We also assessed its performance among groups with differing health status and between subsets of participants stratified according to various sociodemographic attributes. The article concludes by extrapolating from these preliminary findings to a research agenda that would support the deployment of comparable open-ended protocols in other health care settings.

New Contributions

Although several recent articles have explored the meaning of “rigor” in the large-scale elicitation of narratives (Dohan et al., 2016; Grob et al., 2016; Schlesinger, Grob, et al., 2015), this is the first article to provide a comprehensive empirical assessment of how such elicitation protocols perform in practice. Specifically, it documents the extent to which elicited narratives provide a fulsome and representative account of patients’ experiences with clinicians. It also offers the first demonstration that narratives can be elicited from less activated consumers.

Method

For an elicitation protocol to be successfully integrated into existing patient experience surveys, we determined it had to meet two requirements. First, it needed to be short in order to create minimal respondent burden and yield concise responses that clinicians could easily review on a regular basis (Geissler, Friedberg, SteelFisher, & Schneider, 2013; Trigg, 2011). We tested protocols of five to seven questions in length, designed to be completed in roughly 5 minutes, since research suggests that lengthening patient experience surveys to this extent does not erode participation or completion rates (Gallagher & Fowler, 1998). The protocol also had to perform well with both spoken and written responses, since existing patient experience surveys are typically fielded by a combination of mail, telephone, and Internet-based modes (Greene, Speizer, & Wiitala, 2008).

Our research involved multiple stages specifying the evaluative criteria for a protocol; developing methods for measuring those criteria against a “gold standard” account of patient experience; several iterations of testing and refining protocols; and thorough assessment of the final protocol. Because methods for this research initiative are published in greater detail elsewhere, we provide an abbreviated version here (Grob et al., 2016).

Study Participants

Data were collected from a standing Internet panel of more than 50,000 households recruited and maintained by the research firm Gesellschaft für Konsumforschung (GfK). This panel is representative of the American public in terms of sociodemographics, Internet usage prior to joining the panel, and health status (Chang & Krosnick, 2009). A random sample of panelists was invited to participate, and those who agreed were screened for having contact with a clinician in the prior year. A second set of screening questions was then used to recruit three equal-sized strata: Those who reported having a “serious or life-threatening” health event in the previous year; those who reported having a chronic health problem that required regular medical monitoring; and those who had neither of these health problems.

Table 1 presents the characteristics of participants compared to the general American public. Because we oversampled people with recent serious health problems, the sample was older (and therefore more White, given the demographics of the United States) than the population as a whole. Respondents were also more likely to have frequently used health care (10 or more doctor visits in the past year) but were only slightly more likely to have long-established relationships with the clinician about whom they were reporting. Crucial for our analysis of representativeness, more than a third of the sample either (a) had no more than a high school education, (b) was over the age of 65 years, or (c) had a chronic health problem requiring medical monitoring.

Creating a “Gold Standard” for Assessing the Performance of Elicitation Protocols

Because elicitation protocols induce descriptions of health care seen through the patient’s own eyes, their effectiveness can most appropriately be compared to a comprehensive account from that same patient. In our project, this “gold standard” was established by having trained evaluators conduct intensive (45–75 minutes) semistructured telephone interviews with respondents roughly 3 weeks after they completed the elicitation questions. The interviews were open-ended, but incorporated extensive prompts to ensure coverage of 10 domains identified in past studies of patient comments (López, Detz, Ratanawongsa, & Sarkar, 2012). The effectiveness of our protocols was assessed by measuring the congruence or fidelity between the elicited narratives and the paired interview responses.

Defining Evaluative Criteria

Patient narratives have the capacity to augment conventional closed-ended surveys in three ways. The first involves the scope of patient experience information they yield. Narrative accounts often include aspects of care typically omitted from conventional surveys (such as emotive elements of caregiving relationships or aspects of coordinated care that go beyond information exchange) because researchers have difficulty formulating clear questions about these aspects of care and patients have difficulty responding to questions about such aspects using closed-ended scales (Bardach et al., 2015; Greenhalgh & Hurwitz, 1999; Huppertz & Smith, 2014; López et al., 2012; Ranard et al., 2016). Second, narratives can convey a more nuanced representation of how patients evaluate their experiences (Dohan et al., 2016), which are often a mixture of positive and negative events that are muddled together in unclear ways when patients assign overall ratings to their clinicians (Huppertz & Smith, 2014; López et al., 2012). Third, by their nature, narratives have the capacity to convey a

coherent, multidimensional storyline about what happened and how those events unfolded over time that is missing from closed-ended ratings (Bekker et al., 2013; Dohan et al., 2016; Greenhalgh & Hurwitz, 1999; Riiskjaer et al., 2012; Trigg, 2011).

To assess the capacity of elicited narratives to meet their potential in these areas, we developed criteria for the following characteristics: (1) completeness (providing a full picture of the experiences that matter to the patient describing them), (2) balance (accurately reflecting both positive and negative aspects of the patient's experiences), and (3) meaningfulness (conveying a story that is coherent, allowing other patients to assess its relevance for their own future care).

Completeness was measured by parsing patients' interactions with clinicians into 10 domains identified as important to patients in prior research (access, communication, time spent with patient, technical competence, thoroughness, emotional connectedness, coordination of care, shared decision making, practice orientation, and support staff) (Grob et al., 2016; López et al., 2012), then determining how closely experiences reported in the elicited narrative matched those reported in the interviews. Balance was assessed by calculating the ratio of positive to negative statements in the narrative and comparing it to that same ratio from the interview.

Meaningfulness was determined by coding six facets of narrative coherence identified in the literature on narratives related to health (expectations for care, texture of description, emotional and substantive expressivity of the narrative, completeness of storyline, and clear chronology of storyline) (Grob et al., 2016; Reese et al., 2011). Codes for these six facets were averaged to create an overall index of meaningfulness, again compared to the same index constructed for the interview.

The fourth and final criterion for an effective approach to elicitation was the representativeness of the resulting narratives—that is, whether the elicitation protocol was equally effective in capturing experiences from patients across a range of health status and sociodemographic characteristics. Representativeness was assessed by comparing the completeness, balance, and meaningfulness of the average elicited narrative against the properties of narratives elicited from three at-risk groups: Those with no more than a high school education (whose limited literacy or social status might inhibit voice); those with a chronic condition requiring regular medical monitoring (who may feel daunted by their frailty or the complexity of their care); and the elderly (who have above-average social isolation and came of age in an era of more limited medical consumerism). We chose these groups because they are known to be less likely to express their health care experiences to third parties, particularly after experiences that fell short of their expectations (Schlesinger, 2011; Schlesinger, Mitchell, & Elbel, 2002).

Operationalizing Other Key Concepts

Apart from the key outcomes of completeness, balance, meaningfulness, and representativeness (see Table 2), several additional considerations came into play as we evaluated and reported on the performance of the elicitation protocol. The first was a measurement challenge. How can the performance of elicitation protocol be related to the

“salience” of particular aspects of the health care encounters—that is, the perceived importance of different aspects of health care to the patients reporting on their experiences. All else equal, an effective elicitation protocol ought to perform best (according to the four criteria defined above) for those aspects of care that matter most to patients. We measured salience by counting the total number of words that respondents devoted to each domain in the intensive interviews; then examined how salience was related to the effectiveness of the elicitation protocol.

A second consideration was a presentation challenge: How to best way to convey in the context of this article the detailed informational content contained in narratives and to demonstrate how content varied across more and less complete narratives? Numeric measures cannot fully convey the richness and concreteness of description that make narratives valuable and actionable (Cognetta-Rieke & Guney, 2014; Dohan et al., 2016; Friedberg et al., 2011; Riiskjaer et al., 2012; Schlesinger, Grob, et al., 2015; Shaffer, Tomek, & Hulse, 2014; Shaffer & Zikmund-Fisher, 2013; Trigg, 2011; Tsianakas et al., 2012; Winterbottom, Bekker, Conner, & Mooney, 2008). This disjunction seems both ironic and unfortunate in an article such as this one which explores the virtues of narrative feedback. However, since the most fulsome elicited narratives average 300 to 400 words, it is impossible to fully reproduce them here.

To address this limitation, we offer the reader an online appendix presenting four narratives in full, along with commentary comparing each to their far longer matching interviews (averaging 5,000–7,000 words). We selected these examples from different points within the distribution of quantifiable metrics to illustrate the range of elicitations that were more and less successful (according to our specified criteria), though all were collected from respondents who had equally rich stories to tell (i.e., had equally high ratings for completeness and narrative detail based on our coding of their intensive interviews).

Developing and Refining Elicitation Methods Through Multiple Iterations

Our efforts to develop and test elicitation methods were exploratory and iterative. An initial protocol was developed, pilot tested, and fielded for a first round of data collection, then evaluated, revised, and fielded for a second time (Grob et al., 2016). In both stages, the protocol was fielded as part of the Consumer Assessment of Healthcare Providers and Systems Clinician & Group Survey, or CG-CAHPS; the protocol was placed before the closed-ended questions in half of the surveys and after them in the other half.

Data collection for the final protocol and matching interview was conducted in May and June of 2014. The number of matched elicitation surveys and intensive interviews was 51 in the first round of data collection and 48 in the second round. Both protocols were fielded using phone and Internet modalities, with question wording and sequencing identical in each modality.

Narrative fidelity improved significantly between the first and second round. The one exception to this pattern of improvement involved narrative balance, which stayed roughly consistent across both versions of the elicitation protocol: Roughly 80% of patient experiences are reported with a positive valence. This is consistent with the valence of

anecdotes currently posted on websites about clinicians, which average between 65% and 80% positive (Kemp et al., 2016; Lagu, Goff, Hannon, Shatz, & Lindenauer, 2013; López et al., 2012).

The process of refining the protocol yielded a number of insights. We found that the best way to induce a comprehensive and coherent narrative is to create a question sequence that helps respondents construct a narrative arc, beginning with establishing their expectations, then comparing these expectations to their actual experiences, and conveying how they responded when experiences deviated from expectations. Although conventional surveys often use very simply worded questions to reduce ambiguity or the risk of misunderstanding, we found that more complex question wording was necessary to help patients articulate their experiences in ways that can be readily interpreted by diverse audiences of consumers and clinicians. Our final elicitation protocol, presented in full in the appendix to this article, reflects these insights (Grob et al., 2016).

The time required to complete the final elicitation protocol met our criterion for duration: elicitations were completed by phone in 4.7 minutes and by Internet in 5.4 minutes.

Human Subjects Review

This study was reviewed and approved by the Institutional Review Boards (IRB) at Yale University, the University of Wisconsin–Madison, and the RAND Corporation (IRB protocol identifiers, respectively: 1208010703, SE-2012–0582, and 2012–0647-AM01).

Results

Our assessment of the performance of the final version of the elicitation protocol is divided into four stages, corresponding to the performance criteria for evaluating elicited narratives that we introduced above: completeness, balance, meaningfulness, and representativeness.

Completeness of Elicited Narratives

On average, the elicitation protocol yielded descriptions of experiences in slightly less than 4 (3.75) of the 10 measured domains of clinician–patient interaction, compared with an average of 6.35 domains reported in the hour-long intensive interviews. Thus, the elicitation protocol captured about 60% of patient experiences from the past year. The phone and Internet-based elicitation protocols performed equally well in this aspect of fidelity (Figure 1), though with statistically significant differences in performance for care orientation (phone elicitation superior) as well as shared decision making and time availability (web elicitation superior).

However, there was considerable variation in fidelity across the 10 domains of clinician–patient interaction. For each domain, the vertical axis in Figure 1 represents the percent of mentions from the hour-long interview that were also reported in the elicited narrative. In three domains—physician–patient communication, interactions with office staff, and coordination of care—the elicitation protocol captured more than three quarters of the experiences reported in the interviews. In contrast, in three other domains—the physician’s practice orientation (e.g., conventional vs. complementary medicine), thoroughness of

clinician–patient interactions, and shared decision making—elicitations captured only about 15% of the experiences recounted in the interviews.

This variation in fidelity across domains is related, at least in part, to differences in the salience of these domains for patients (Grob et al., 2016). Combined, the three high-fidelity domains represented 38% of the total word count in the intensive interviews, while the three low-fidelity domains represented only 11% of the total word count. In other words, the elicitation protocol worked best (most reliably) for aspects of health care that are most readily recalled with minimal prompting.

Balance in Elicited Narratives

Elicited narratives yielded accounts of physician–patient interactions that were 80% positive. As noted earlier, this closely matches the balance observed in more anecdotal accounts, suggesting that the anecdotes currently found on the Internet are not primarily coming from patients aggrieved about their health care experiences.

How does this balance compare to that observed in the intensive interviews? Because this is an important aspect of performance for any elicitation protocol, we measured it in four distinct ways. The first two were derived from a content analysis of the narratives and interviews: (1) the balance of positive versus negative events or aspects of care reported by the respondent and (2) the balance of the total words used to depict these positive and negative experiences. The other two measures were designed to capture how the elicited narrative would “feel” to readers (Reese et al., 2011). After reading each elicited narrative and intensive interview, the coders of the transcripts were asked to make their own subjective assessment of how positively the respondent regarded their interactions with both their clinician and their interactions with that clinician’s office personnel, since assessments of staff were often more critical than those of clinicians. Inter-coder reliability was acceptable for each of these subjective codes ($\kappa > .70$ in all cases).

Our measure of balance was constructed as the ratio of the positive versus negative balance (for each of the four metrics) for the elicited narrative divided by the positive versus negative balance for the matched interview. Given this formula, a ratio equal to one means that the interview and elicited narrative had exactly the same balance of positive versus negative assessments. If the ratio exceeds one, the narrative is positively biased (relative to the interview); if the ratio is less than one, the narrative is negatively biased.

The results (Figure 2) suggest that the elicited narratives offer a reliably balanced representation of patient experience. The two measures based on content analysis had the narratives within 10% of the positive/negative mix depicted in the intensive interviews. The measures based on the subjective “feel” of the narratives displayed a modest positive bias, particularly in the representation of interactions with office staff (but this positive skew was still less than 20% of the mean). Overall, telephone administration of the protocol yielded a slightly more positive representation of patient experiences than web administration. None of the differences across elicitation modes were statistically significant.

Meaningfulness

Meaningfulness was primarily assessed through a composite measure of average coder scores for six attributes of the narrative (Grob et al., 2016). We also explored patterns in each of the component measures, two of which are reported separately in Figure 3 to help explain patterns in the composite.

The elicitation protocol performed about as well in generating meaningful narratives as it did in generating complete narratives: The average meaningfulness score for the storyline from the 5-minute elicitations was about 60% of the score for the corresponding hour-long interviews. For this aspect of the open-ended accounts, the phone mode consistently outperformed the Internet mode. This was true even though there was no difference in the wording or sequencing of the questions, and the telephone interviewers for the survey were carefully instructed to limit their interactions with respondents to simply reading them the protocol questions.

Analysis of the six component measures reveals that narratives obtained by phone and Internet were similar with regard to most of these measures. We illustrate this in Figure 3 using one of these measures: the completeness of the storyline—that is, the extent to which the elicited narrative conveyed a logical sequence of interactions with clinicians. The one aspect of narratives for which the phone and Internet narratives differed markedly (and statistically significantly) involved texture, that is, the extent of detail provided by patients when describing interactions: phone elicitations had almost twice as much detail. Why the phone mode yielded greater detail is not evident; we suspect that it may simply have been easier to recount details verbally rather than in writing.

Illustrating Attributes of Elicited Narratives

As we noted above, the elicited narratives varied in their detail and complexity. However, even some of the least successful elicited narratives offered actionable clues for improving clinicians' practices. The most complete elicited narratives were rich in detail, describing as many as a half dozen distinct aspects of the health care experience that might be improved in the future. Although space here is too limited to pursue the concept of "actionability"—that is, data that can be meaningfully harnessed for quality improvement—in a complete manner, the four narratives presented in the online appendix illustrate both the variability and potential actionability of collected data (see Supplemental Digital Content, available in the online version of the article).

Representativeness

Representativeness is assessed in terms of the *relative* richness of narratives available from different subsets of respondents. We compared elicitations for the average respondent to those from three subgroups at risk for weaker voice: the less educated, the most medically frail, and the most socially isolated. In addition to the quantifiable metrics reported above, we also assessed the word count on the elicited narrative as a percent of the word count on the matching interview.

As shown in Figure 4, the overall pattern seems clear: none of these three groups appears to have been particularly disadvantaged in conveying their experiences through elicited narratives (i.e., none of the cross-group differences are statistically significant for any of the outcome measures). Indeed, the elderly seem to convey narratives that are consistently richer than average (though again, these differences are not statistically significant), perhaps because of the greater salience of health issues in their lives. This above-average performance is somewhat offset by a slight positive bias in the elicited narratives from older Americans.

Discussion

The findings presented here demonstrate that a 5-minute, five-question elicitation protocol incorporated into a conventional patient experience survey (fielded either by phone or in writing via the web) can yield narratives that accurately replicate the balance of positive and negative experiences that patients have with their clinicians and capture with reasonable fidelity the depth and nuances of patient experiences. Though no short, structured narrative can match the richness of a semistructured intensive interview, our findings illustrate the feasibility of designing elicitation protocols that are sufficiently concise to be incorporated into large-scale surveys, while still yielding substantial, concrete information about patient experience. They also show that a protocol for eliciting patient narratives can be equally effective in subsets of patients that could potentially be least willing or able to express their experiences.

Study Limitations and Responses

The findings presented here need to be considered in light of certain limitations of our methods. First and foremost, there are no clear criteria for determining when elicited narratives have yielded sufficient information that they adequately portray patient experience. To be sure, clinicians currently have limited ability to determine why their patients are evaluating them positively or negatively, other than their scores on broad survey composites (e.g., communication, access, coordination of care or staffing). So one might anticipate that any approach that conveys a substantial portion of that experience in concrete detail (as the elicitation protocol tested here clearly does) would be seen as valuable to clinicians—and past evidence suggests that clinicians do in fact find patients' narrative accounts to be useful for quality improvement (Geissler et al., 2013). Nonetheless, it is certainly possible that other elicitation protocols might yield even more fulsome and useful accounts; discovering whether this is so remains a matter for future research.

There may also be alternative ways to operationalize an objective gold standard for assessing narrative content. For example, we measured the completeness and meaningfulness of the elicited narratives by matching narrative and interview content across multiple aspects of care and dimensions of narrative coherence. The ratios presented in Figures 1, 3, and 4 presume that each “match” between narrative and interview content is equally valued, whether or not the reported aspects of care were equally salient to the patients in question or equally important for enhancing quality. Realistically, these ratios should be viewed as heuristics for the performance of narrative elicitation, rather than as precise measures of

some continuously measured attribute of narrative content or coherence. Future research might explore alternative standards of comparison. These could include comparing narrative and interview content in terms of the aspects of care that patients deem most essential or that practice administrators consider most actionable for quality improvement.

Moreover, it seems likely that narratives elicited as part of large-scale patient experience surveys, however, effective the elicitation protocol, will be more useful in identifying problem areas that need to be addressed than in offering definitive solutions. That is because the elicited narratives often identify *what* went wrong but not always *why*, causal attributions being challenging for many patients (Rosenthal & Schlesinger, 2002). When problem areas are identified by elicited narratives, it may then become essential to use other techniques to probe more deeply into those aspects of patient experience, including more intensive forms of patient interviews or focus groups.

Although comparing elicited narratives to matched interviews offers one useful take on the effectiveness of the elicitation protocol, it does not assess what narratives contribute compared with close-ended survey questions about patient experience. There has been relatively little research about this in ambulatory care settings, though the limited evidence suggests that narrative content helps to explain patients' overall assessment of their clinicians above and beyond their responses to conventional CAHPS survey questions (Martino et al., 2017). There has, however, been considerable research on narratives for inpatient care. Here too narrative comments have been found to help explain patients' assessment of hospital care above and beyond responses on H-CAHPS closed-ended questions (Kemp et al., 2016). In addition, there is extensive evidence that narratives convey information about many domains of care that are not covered by conventional patient experience surveys (Bardach et al., 2015; Gallan, Girju, & Girju, 2017; Huppertz & Smith, 2014; Ranard et al., 2016), often in ways that are considered actionable by clinicians (Riiskjaer et al., 2012; Tsianakas et al., 2012).

Finally, testing the effectiveness of the elicitation protocol using an Internet-based panel may yield results different from a test involving the general population of patients. In particular, patients who are not literate or not proficient in written English might respond differently to requests to recount their experiences. One might expect phone-based elicitations to be less sensitive to these constraints on fulsome responses; it is helpful to observe from our findings that responses via web elicitations were at least as complete as those collected over the phone. But phone elicitations may require additional resources to transcribe answers, unless the reliability of voice recognition software improves.

Study Extensions

These results hold promise for more effectively harnessing the potential of patient narratives to inform consumers, help clinicians respond to patient needs, and enhance policy makers' capacity to monitor health system performance. At the same time, our work represents just the first steps along an extended path of implementation. Three sets of challenges will need to be addressed to continue making progress.

The first challenge is expanding the scope of narrative elicitation. As with conventional patient experience surveys, elicitation protocols would need to be adapted to the circumstances of particular health care settings. We tested an elicitation protocol focused on patients' experiences with clinical care in ambulatory settings. We do not yet know what kinds of protocols would be successful for hospitals, health plans, or various kinds of specialty care.

It may also be useful to develop specialized protocols that directly inquire about particular aspects of care. For example, coordination between primary care clinicians and specialists is a core objective of patient-centered medical homes (Roseman, Osborne-Stafsnes, Amy, Boslaugh, & Slate-Miller, 2013) but was of sufficiently low salience in our study that comments related to coordination were mentioned by only about 10% of respondents (Grob et al., 2016), making it difficult to reliably track these aspects of care using a generalized protocol.

A second set of challenges involves effective use of the narratives collected through elicitation protocols. The details embedded in narratives can provide clinicians with vital insights about practice improvement, but it may be difficult for them to interpret a substantial ongoing stream of narratives without sufficient time and appropriate administrative support (Schlesinger, Shaller, et al., 2015). This support could include forms of natural language processing to help identify patterns in the narrative responses. But we anticipate the need to apply these programs carefully, to avoid having processed feedback substitute for clinicians' more nuanced reading of their patients' experiences, described in their own words.

The concreteness of patient narratives can also help other consumers identify clinicians whose practice styles best match their needs and preferences, but that same detail can also distract consumers from other useful metrics of quality (Kanouse et al., 2016). To overcome these challenges, it will be essential to develop methods to integrate narrative content into quality reports effectively. However this is done, strategies for reporting narratives to clinicians and consumers must be designed and tested with the same rigor applied here to the development of the elicitation protocol.

Finally, integrating open-ended narratives into conventional patient experience surveys poses challenges related to the logistics of survey administration. Narrative comments are most readily collected and processed for surveys that are collected online, yet Internet administration remains one of the least common modalities for fielding patient experience surveys (Shaller, Dodson, Ricketts, Li, & Yount, 2015). More time and consideration will also have to be given to how open-ended questions can best be deployed to enhance survey participation, despite the longer length of the instruments.

Embracing a conception of patient-centered health care fully enriched by patient-reported information calls for a commitment to collecting narratives that are truly representative of diverse patients' experiences with different providers. Narratives are not simply the icing on the cake of patient-reported outcomes (Dohan et al., 2016; Trigg, 2011). In many cases, they will provide the crucial substantive information required to make medical practice more

responsive to patients' needs and preferences. So too, narratives can play a crucial and growing role in making consumer choice among providers truly informed.

To realize this potential in the face of an ever-changing health care landscape requires a further commitment to a dynamic conception of rigor—to constantly revising and improving the collection and presentation of narratives. That commitment must be embraced with the same enthusiasm currently devoted to bringing other new technologies into medical settings. The findings presented here should thus be seen as simply a starting point. They represent a proof of concept: evidence that rigor is possible in the collection of narrative accounts. We hope they serve as the precursor to a far broader, sustained commitment to narrative elicitation and reporting across the health care system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project on which this article is based was facilitated by helpful feedback from a variety of sources, including the broader CAHPS research team (principal investigators Paul Cleary, Susan Edgman-Levitan, Ron Hays and Marc Elliott); our project officers at AHRQ (Chris Crofton and Caren Ginsberg); participants in seminars at Baystate Medical Center, the University of Wisconsin–Madison, and Oxford University; and attendees at several web-based conferences with staff and leadership from AHRQ and CMS.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by cooperative agreements between AHRQ and Yale (1U18HS016978) and RAND (2U18HS016980–06) as well as a grant from AHRQ to Yale University (1R21HS021858).

Appendix

The Core Protocol

Q1. What are the most important things that you look for in a health care provider and his or her staff?

Q2. When you think about the things that are most important to you, how do your provider and his or her staff measure up?

Q3. Now we'd like to focus on anything that has gone well in your experiences with your provider and his or her staff over the past 12 months. Please explain what happened, how it happened, and how it felt to you.

Q4. Next we'd like to focus on any experiences with your provider and his or her staff that you wish had gone differently over the past 12 months. Please explain what happened, how it happened, and how it felt to you.

Q5. Please describe how you and your provider relate to and interact with each other.

Supplemental Questions

[These were asked as the sixth and seventh questions in sequence. Because they elicited infrequent responses, they were deemed more appropriate as supplemental questions]

S1. Think back to those experiences you wish had gone differently. What, if anything, did you or your provider do about the situation and how did it turn out? For example, did you talk to your doctor about the issue, talk to your family or friends, look for another doctor, take any other form of action, or do nothing at all?

S2. Think back to those experiences that went well. What, if anything, did you do in response to these positive experiences? For example, did you talk to your doctor about them, talk to your family or friends about them, recommend this doctor to other people, take any other form of action, or do nothing at all?

References

- Bardach NS, Lyndon A, Asteria-Peñaloza R, Goldman LE, Lin GA, & Dudley RA (2015). From the closest observers of patient care: A thematic analysis of online narrative reviews of hospitals. *BMJ Quality & Safety*, 25, 889–897. doi:10.1136/bmjqs-2015-004515
- Bekker HL, Winterbottom AE, Butow P, Dillard AJ, Feldman-Stewart D, Fowler FJ, ... Volk RJ (2013). Do personal stories make patient decision aids more effective? A critical review of theory and evidence. *BMC Medical Informatics and Decision Making*, 13(Suppl. 2), S9. [PubMed: 24625283]
- Chang L, & Kroshnick JA (2009). National surveys via RDD telephone interviewing versus the internet: Comparing simple representativeness and response quality. *Public Opinion Quarterly*, 73, 641–678.
- Cognetta-Rieke C, & Guney S (2014). Analytical insights from patient narratives: The next step for better patient experience. *Journal of Patient Experience*, 1(1), 22–24. [PubMed: 28725805]
- Dohan D, Garrett SB, Rendle KA, Halley M, & Abramson C (2016). The importance of integrating narrative into health care decision making. *Health Affairs*, 35, 720–725. [PubMed: 27044974]
- Friedberg MW, SteelFisher GK, Karp M, & Schneider EC (2011). Physician groups' use of data from patient experience surveys. *Journal of General Internal Medicine*, 26, 498–504. [PubMed: 21161419]
- Gallagher P, & Fowler FJ (1998). Size doesn't matter: Response rates of Medicaid enrollees to questionnaires of varying lengths. Paper presented at American Association for Public Opinions Research Meeting, St. Louis, MO.
- Gallan AS, Girju M, & Girju R (2017). Perfect ratings with negative comments: Learning from contradictory patient survey responses. *Patient Experience Journal*, 4(3), 15–28.
- Geissler KH, Friedberg MW, SteelFisher GK, & Schneider EC (2013). Motivators and barriers to using patient experience reports for performance improvement. *Medical Care Research and Review*, 70, 621–635. [PubMed: 23877957]
- Greaves F, Millett C, & Nuki P (2014). England's experience incorporating "anecdotal" reports from consumers into their national reporting system: Lessons for the United States of what to do or not to do? *Medical Care Research and Review*, 71(5 Suppl), 65S–80S. [PubMed: 24836765]
- Greene J, Speizer H, & Wiitala W (2008). Telephone and web: Mixed-mode challenge. *Health Services Research*, 43(1 Pt 1), 230–248. [PubMed: 18211527]
- Greenhalgh T, & Hurwitz B (1999). Narrative based medicine: Why study narrative? *British Medical Journal*, 318, 48–50. [PubMed: 9872892]
- Grob R, Schlesinger M, Parker AM, Shaller D, Barre LR, Martino SC, ... Cerully JL (2016). Breaking narrative ground: Innovative methods for rigorously eliciting and assessing patient narratives. *Health Services Research*, 51(Suppl. 2), 1248–1272. [PubMed: 27126144]

- Hao H (2015). The development of online doctor reviews in China: An analysis of the largest online doctor review website in China. *Journal of Medical Internet Research*, 17, e134. doi:10.2196/jmir.4365 [PubMed: 26032933]
- Holliday AM, Kachalia A, Meyer GS, & Sequist TD (2017). Physician and patient views on public physician rating websites: A cross-sectional study. *Journal of General Internal Medicine*, 32, 626–631. [PubMed: 28150098]
- Huppertz JW, & Smith R (2014). The value of patients' handwritten comments on HCAHPS surveys. *Journal of Healthcare Management*, 59, 31–48. [PubMed: 24611424]
- Kanouse DE, Schlesinger M, Shaller D, Martino SC, & Rybowski L (2016). How patient comments affect consumers' use of physician performance measures. *Medical Care*, 54, 24–31. [PubMed: 26551765]
- Kemp K, Warren S, Chan N, McCormack B, Santana M, & Quan H (2016). Qualitative complaints and their relation to overall hospital rating using an H-CAHPS-derived instrument. *BMJ Quality & Safety*, 25, 770–777. doi:10.1136/bmjqs-2015-004371
- Lagu T, Goff SL, Hannon NS, Shatz A, & Lindenauer PK (2013). A mixed-methods analysis of patient reviews of hospital care in England: Implications for public reporting of health care quality data in the United States. *Joint Commission Journal on Quality and Patient Safety*, 39, 7–15. [PubMed: 23367647]
- Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, & Lindenauer PK (2017). Website characteristics and physician reviews on commercial physician-rating websites. *JAMA Journal of the American Medical Association*, 317, 766–768. [PubMed: 28241346]
- López A, Detz A, Ratanawongsa N, & Sarkar U (2012). What patients say about their doctors online: A qualitative content analysis. *Journal of General Internal Medicine*, 27, 685–692. [PubMed: 22215270]
- Martino SC, Shaller D, Schlesinger M, Parker AM, Rybowski L, Grob R, ... Finucane ML (2017). CAHPS and comments: How closed-ended survey questions and narrative accounts interact in the assessment of patient experience. *Journal of Patient Experience*, 4(1), 37–45. [PubMed: 28725858]
- Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, ... Merchant RM (2016). Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. *Health Affairs*, 35, 697–705. [PubMed: 27044971]
- Reese E, Haden CA, Baker-Ward L, Bauer P, Fivush R, & Ornstein PA (2011). Coherence of personal narratives across the lifespan: A multidimensional model and coding method. *Journal of Cognition and Development*, 12, 424–462. [PubMed: 22754399]
- Riiskjaer E, Ammentorp J, & Kofoed PE (2012). The value of open-ended questions in surveys on patient experience: Number of comments and perceived usefulness from a hospital perspective. *International Journal for Quality in Health Care*, 24, 509–516. [PubMed: 22833616]
- Roseman D, Osborne-Stafsnes J, Amy CH, Boslaugh S, & Slate-Miller K (2013). Early lessons from four “aligning forces for quality” communities bolster the case for patient-centered care. *Health Affairs*, 32, 232–241. [PubMed: 23381515]
- Rosenthal M, & Schlesinger M (2002). Not afraid to blame: The neglected role of blame attribution in medical consumerism and some implications for health policy. *Milbank Quarterly*, 80, 41–95. [PubMed: 11933793]
- Schlesinger M (2011). The canary in the gemeinschaft: The public voice of patients as a means of enhancing health system performance In Hoffman B, Tomes N, Grob R & Schlesinger M (Eds.), *Patients as policy actors: A century of changing markets and missions* (pp. 148–176). New Brunswick, NJ: Rutgers University Press.
- Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, ... Rybowski L (2015). Taking patients' narratives about clinicians from anecdote to science. *New England Journal of Medicine*, 373, 675–679. [PubMed: 26267629]
- Schlesinger M, Mitchell S, & Elbel B (2002). Voices unheard: Barriers to expressing dissatisfaction to health plans. *Milbank Quarterly*, 80, 709–755. [PubMed: 12532645]
- Schlesinger M, Shaller D, & Grob R (2015). Using patient-reported information to improve clinical practice. *Health Services Research*, 50(Suppl. 2), 2116–2154. [PubMed: 26573890]

- Shaffer VA, Tomek S, & Hulsey L (2014). The effect of narrative information in a publicly available patient decision aid for early-stage breast cancer. *Health Communication*, 29, 64–73. [PubMed: 23384155]
- Shaffer VA, & Zikmund-Fisher BJ (2013). All stories are not alike: A purpose-, content-, and valence-based taxonomy of patient narratives in decision aids. *Medical Decision Making*, 33, 4–13. [PubMed: 23065418]
- Shaller D, Dodson T, Ricketts J, Li S, & Yount N (2015). CAHPS Clinician and Group Survey Database 2014 Chartbook: What patients say about their health care providers and medical practices (AHRQ Publication No. 15-CAHPS002-EF). Retrieved from <https://www.cahpsdatabase.ahrq.gov/files/2014CAHPSClinicianGroup.htm>
- Sick B, & Abraham JM (2011). Seek and ye shall find: Consumer search for objective health care cost and quality information. *American Journal of Medical Quality*, 26, 433–440. [PubMed: 21918015]
- Sun LH (2015, 6 3). How hospitals hope to boost ratings on Yelp, HealthGrades, ZocDoc and Vitals. The Washington Post. Retrieved from https://www.washingtonpost.com/national/health-science/sites-like-yelp-can-be-tough-but-hospitals-embrace-online-reviews/2015/06/03/a07a68b6-fe63-11e4-805c-c3f407e5a9e9_story.html?utm_term=.060161e894e5
- Tompson TN, Agiesta J, Alvarez E, Benz JK, Malato D, Reimer R, ... Willcoxon N (2014). Finding quality doctors: How Americans evaluate provider quality in the United States. Retrieved from <http://www.apnorc.org/PDFs/Finding%20Quality%20Doctors/Finding%20Quality%20Doctors%20Research%20Highlights.pdf>
- Trigg L (2011). Patients' opinions of health care providers for supporting choice and quality improvement. *Journal of Health Services Research & Policy*, 16, 102–107. [PubMed: 20870690]
- Tsianakas V, Maben J, Wiseman T, Robert G, Richardson A, Madden P, ... Davies EA (2012). Using patients' experiences to identify priorities for quality improvement in breast cancer care: Patient narratives, surveys or both? *BMC Health Services Research*, 12, 271. [PubMed: 22913525]
- Winterbottom A, Bekker HL, Conner M, & Mooney A (2008). Does narrative information bias individual's decision making? A systematic review. *Social Science & Medicine*, 67, 2079–2088. [PubMed: 18951673]

Figure 1.

Completeness in scope of reported experiences.

Note. Calculated by authors, data collected from GfK (Gesellschaft für Konsumforschung) Knowledge Panel.

Figure 2.

Balance of positive versus negative assessments.

Note. Calculated by authors, data collected from GfK (Gesellschaft für Konsumforschung) Knowledge Panel.

Figure 3.

Attributes of the narrative in elicited comments.

Note. Calculated by authors, data collected from GfK (Gesellschaft für Konsumforschung) Knowledge Panel.

Figure 4.

Representativeness: Elicitation performance by subsets of respondents.

Note. Calculated by authors, data collected from GfK (Gesellschaft für Konsumforschung) Knowledge Panel.

Table 1.

Elicitation Sample Compared With U.S. Population.

Characteristic	Elicitation sample (%)	U.S. population (%)
Sociodemographics ^a		
Age (years)		
<30	11.0	21.5
30–44	17.2	25.5
45–60	28.2	27.1
>60	43.6	25.8
Race/ethnicity		
White	78.0	66.0
African American	10.3	11.6
Latino	7.0	15.0
Other	4.8	7.5
Education		
High school or less	38.9	42.2
Some college	30.0	28.9
College graduate	31.1	28.9
Health status and utilization ^b		
Chronic health problems		
Yes	36.6	49.8
No	63.4	50.2
Physician visits in previous year		
1	21.2	31
2–3	42.8	43
4–9	28.6	22
>9	6.6	4
Time with current physician		
1 year	23.5	37
2–3 years	20.1	19
3–4 years	16.8	12
5+ years	38.8	32

Source. Elicitation sample characteristics calculated by authors, data collected from GfK (Gesellschaft für Konsumforschung) Knowledge Panel.

^aU.S. population statistics from the Current Population Survey, 2014.

^bHealth care utilization statistics from the Consumer Assessment of Healthcare Providers and Systems benchmarking database.

Table 2.

Key Attributes for Evaluating Elicited Patient Narrative.

Evaluative attribute	Operational definition
Completeness	Elicited narrative includes experiences in the same domains (out of 10 possible) that the patient experienced and reported in their interview
Balance	Elicited narrative includes the same ratio of positive/negative experiences as in the interview
Meaningfulness	Elicited narrative meets the same six criteria of narrative coherence as did the interview
Representativeness	Elicitation protocol performs equally well (based on the three criteria identified above) for subsets of patients who have been historically at risk for impaired voice: older, less educated, more seriously ill