Rapid Sense Making: A Feasible, Efficient Approach for Analyzing Large Data Sets of Open-Ended Comments

Rebecca S. Etz1, Martha M. Gonzalez1, Aimee R. Eden2, and Jodi Winship3

Abstract
This article shares the problem-solving process and resultant rapid sensemaking methodology created by an interdisciplinary research team faced with qualitative “big data.” Confronted with a data set of over half a million free text comments, within an existing data set of 320,500 surveys, our team developed a process to structure the naturally occurring variability within the data, to identify and isolate meaningful analytic units, and to group subsets of our data amenable to automated coding using a template-based process. This allowed a significant portion of the data to be rapidly assessed while still preserving the ability to explore the more complex free text comments with a grounded theory informed emergent process. In this discussion, we focus on strategies useful to other teams interested in fielding open-ended questions as part of large survey efforts and incorporating those findings as part of an integrated analysis.

Keywords
methods, qualitative, mixed methods

What Is Already Known?
Open-ended questions are commonly used in surveys, but methods to analyze the free text comments generated as part of large survey efforts are not well-developed. A feasible and efficient method is needed to analyze responses generated from open-ended questions.

What Does This Paper Add?
This article articulates a rapid sensemaking (RSM) approach that includes three easy-to-use, feasible, and efficient protocols to enable identification of data patterns and significant outliers as well as easy integration of qualitative and quantitative survey findings. These protocols allow analysts to (1) reduce noise by eliminating data that are not analytically useful, (2) identify and structure data variability by type and complexity, and (3) assign coding strategies appropriate to identified data subsets.

Background
The inclusion of open-ended questions in the structured surveys is a useful technique in health services research. Open-ended questions allow researchers to probe for potentially hidden nuances within survey responses, to solicit additional comments not confined by predetermined categories, and to expose the unexpected, including responses that might challenge the assumptions upon which the structure of the survey is based (Marcinowicz, Chlabicz, & Grebowski, 2007; Riisjaer, Ammentorp, & Kofoed, 2012).

However, analyzing the free text comments generated by open-ended questions as part of large survey efforts can require significant time and resources (O’Cathain & Thomas, 2004). For this reason, and despite the potential benefits, health service researchers who field large surveys often opt out of

1 Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, VA, USA
2 Research Department, American Board of Family Medicine, Lexington, KY, USA
3 Department of Health Behavior and Policy, Virginia Commonwealth University, Richmond, VA, USA

Corresponding Author:
Rebecca S. Etz, Department of Family Medicine and Population Health, Virginia Commonwealth University, 830 East Main Street, PO Box 980101, Richmond, VA 23298, USA.
Email: rebecca.etz@vcuhealth.org

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
including open-ended questions. Those who do seldom find a satisfactory way for either analyzing or incorporating the free text comments they receive. As a result, the impact of survey-based free text comments, especially in large data sets, is often underdeveloped or lost (Rich, Chojenta, & Loxton, 2013).

Required is a clearly articulated and efficient approach for making sense of large data sets of free text comments to enable their integration with other study findings. Methodological literature about analyzing free text comments often differentiates between manual human coding and semiautomated computer-assisted textual analysis (Jackson & Trochim, 2002; Roberts et al., 2014; Wiedemann, 2013). Previous studies of free text data from patient experience surveys containing 300–5,000 comments have described this analytic process as a largely manual content analysis, or manual thematic coding process, that entailed reading all comments received (Bankauskaite & Osmo, 2003; Bracher, Corner, & Wagland, 2016; Bracher, Wagland, & Corner, 2014; Cunningham & Wells, 2017; Morin, et al., 2015; Richards, Campbell, Walshaw, Dickens, & Greco, 2009; Wiedemann, 2013). However, data sets of greater than 5,000 comments make that a formidable and burdensome task. For larger qualitative data sets, computer assistance makes the task more feasible. Approaches that incorporate natural language processing and other types of qualitative data analysis software have been described to varying levels of detail in the literature (MacRae et al., 2015; Maramba et al., 2015; Namey, Guest, Thairu, & Johnson, 2008; Nasukawa & Yi, 2003; Wiedemann, 2013). Natural language processing algorithms seem best suited for analyzing sentiment or for application to a specific type of content such as data from electronic health records (MacRae et al., 2015; Nasukawa & Yi, 2003). Other text-mining tools provide greater flexibility but often simply assist with management of manually read and coded data (Namey et al., 2008; Wiedemann, 2013).

The existing literature is lacking concrete steps to address certain predictable and common challenges that naturally occur within large data sets of short but unstructured comments; steps that could allow for computer-assisted sorting into subsets and incorporation of some portion of automated coding. These common challenges include (1) a significant amount of noise caused by random keystrokes or nonmeaningful comments, (2) a lack of structure that prevents logorhythmic decision-making or automated functions, and (3) the ability to apply a diverse but integrated set of coherent coding strategies. To address these challenges, our team used a large, mixed data set from a survey completed by over 320,500 respondents to develop a rapid, feasible, and efficient method for making sense of free text survey–based comments. The RSM approach that resulted draws upon existing analytic tools and methods to address concrete problems faced when attempting to analyze very large quantities of qualitative data.

Method

All data management and analysis strategies described in the RSM method were first tested using a purposefully selected 10% subsample of our data. We refer to this as our test environment. Which data to include for an appropriate test environment will vary by study? We chose the content of our test environment to maximize comment diversity in order to test the flexibility and utility of our management and analysis strategies. This meant choosing to include comments from a range of question styles within the survey as well as survey domain areas. During method development, we carefully documented the trial and error process by which each challenge was solved. When an automated strategy applied in the test environment appeared useful, we repeated the process manually with a trained coding team in order to test the reliability of the outcome. Once a solution appeared salient, it was tested de novo in a second 10% subset of our data, randomly selected. If comments were subject to removal from the full data set as the result of an automated process, the entirety of those comments were inspected to check for potential data loss that might have implications for later analyses. Comments that were removed through an automated process and yet found on manual examination to be potentially useful were considered data loss. We conservatively established a baseline of 5% data loss or 5% disagreement between an automated process and a manual process (therefore providing 95% confidence) as acceptable and unlikely to influence future analyses.

Data Management

The development of our RSM approach was organized around the need to address three common yet resilient challenges to the analysis of large data sets of survey-based free text comments: noise, variability, and the need to incorporate protocol-driven automation. Our methodology was therefore designed around problem-solving techniques focused on these three challenges.

After familiarizing ourselves with the sample data set (described below), our first interest was in cleaning, or reducing, the size of our data (Bracher et al., 2014; Cunningham & Wells, 2017). We eliminated noise by automating the removal of comments that were not meaningful (e.g., “#%$” or “...”), and meaningful comments that were not analytically useful (e.g., “n/a” or “no comment”). Once cleaned, we divided the remaining data into subsets based on the structural characteristics of comment length and match of comment content to the question asked (see Figure 1). Comments identified as directly responsive, or appropriate, to the question asked were defined as contextually significant. However, some comments were responsive to the focus of the survey but not to the question asked. Comments identified as not directly responsive to the question but topically appropriate to the survey were defined as conceptually significant.

We then considered the internal characteristics of each data subset in order to maximize automated coding solutions described below. Internal characteristics were defined by the complexity of free text comments. We identified three levels of complexity: simple, compound, or complex. Comments are deemed simple when they are relatively short and convey a single, easily identified concept or idea. Compound comments...
are like two simple comments combined—they are slightly longer but have two or more easily identified concepts or ideas. Simple and compound comments are amenable to the automated coding strategies described below. Complex comments convey nuanced ideas or use language not easily isolated through automated key word coding approaches.

Data Analysis

Coding strategies were matched to data subsets based on both structural characteristics and internal characteristics (level of comment complexity). Automated coding based on research team identified key words was used for both contextually and conceptually significant comments (structural characteristics) that were either simple or compound (internal characteristics). Complex comments required a manual approach that incorporated a combination of template-based coding and grounded theory (Crabtree & Miller, 1999).

Dividing our data set into subsets appropriate for automated coding was facilitated by the use of macros in a tabular software environment. While we used Microsoft Excel for this purpose, any similar tabular software would be appropriate. Macros are mini programs that automate predictable, repetitive sets of keystrokes or functions. Instructions regarding how to create simple macros, such as the ones we used, are easily found online and among tech chat rooms. We used Visual Basic—a free and easy to use programming language that functions in any Microsoft environment—to create macros that identified groups of data based on the length of entry or length of entry combined with potential key words. Some key words were selected in advance by the research team, based on likely responses to survey questions and knowledge of the expert knowledge of the survey domain content. This was supplemented by iterative visual scans of “simple comments” within the text environment to identify common word choices and response patterns. Creating a macro to temporarily isolate comments of minimal length, in our case fewer than 21 characters, isolates comments most likely to reveal additionally useful key words.

Once coded, each research team will need to determine the best process for understanding the significance of coding patterns, frequencies, and co-occurrences. Our sample data set was designed to pair questions that solicited scaled numeric responses with opportunities for free text comments. Once coded, this format allowed for qualitative findings to inform quantitative findings and vice versa. For example, our surveys solicited feedback regarding 16 learning modules (sample data set described in detail below). While the numeric ratings can answer the question of which modules are preferred, or which modules rank higher among learners than others, qualitative findings can answer the question of why those modules are preferred. There are also many times when quantitative findings are ambiguous, for instance, when many parts of a learning activity are rated poorly and yet the overall activity is rated highly. Qualitative findings specific to the low-rated questions can be compared with qualitative findings specific to the overall rating question to explain the discrepancy.

The same ability to compare coding trends with numeric rating trends can be used to understand the relative significance of respondent comments. A frequent and negative trend regarding learner experience in qualitative comments may lead to the impression that the activity is poorly designed. However, if when compared to numeric ratings, these findings are not matched with lower numeric scores, it is possible that the sentiment, while frequently expressed and in need of attention, is not defining of the learner’s experience. Automated coding based on free text comments paired with scaled numeric ratings makes it easy to allow qualitative and quantitative findings to inform the significance of each other. Additionally, being able to isolate those comments not amenable to automated coding is the best way to identify data potentially able to generate new and unexpected findings, not predetermined by survey domain content or survey design.

Figure 1. Flow diagram for subgroup assignment and associated coding method.
Our Sample: Description of the Data Set Used for Method Development

Before sharing our results, it is important for the reader to know something about the data set used to develop our approach. Since 2003, board certification of family physicians by the American Board of Family Medicine (ABFM) has included a set of continuous learning activities. One such activity is the self-assessment module (SAM). SAMs are structured learning activities centered around an independent review of topically identified reading materials (e.g., best practice and evidence regarding asthma treatment). Each review of reading materials is followed by two structured learning activities: a knowledge assessment component (quiz) and a skills assessment component (virtually based clinical simulation).

Each time a learner participates in an SAM, they are asked to complete a survey based on their experience. Feedback surveys include structured questions employing a “Likert-type” scale followed by the opportunity for a free text comment. The length of ABFM feedback surveys varied by topical focus. In general, surveys included 12–20 scaled questions, each paired with the opportunity for free text comment. While each survey was tailored to the content of a specific module, six general survey domains were consistently represented. These included respondent reflections on appropriateness of SAM content, skill acquisition, learning objectives, learning experience, user interface, and overall impressions.

By April 2013, the ABFM had accrued over 320,500 SAM feedback surveys across 16 topic modules. Although the ABFM reviewed free text comments on a periodic basis, in 2013, the board asked our team to conduct a robust analysis of what learners had said in their comments over time. Deidentified data were shared with our team as 16 Excel workbooks, one for each SAM topic. These data contained a potential 5.2 million free text comments—what our team refers to as the codeable universe (this and other key terms defined in Table 1). That codeable universe is the setting in which we developed our approach.

The RSM method is described in greater detail below, through explanation of our problem-solving approach and through grounding in the discussion of how we handled our sample data. To facilitate review of our Results section, we briefly outline the RSM method here as follows (see also Figure 1):

1. Identify and remove noise within the data set.
2. Identify and classify remaining data as (a) meaningful but not useful, (b) meaningful but of limited use, or (c) meaningful and useful.
   a. Data fitting category (b) will not always be present.
   b. All following steps are applied to data fitting categories (b) and (c) only.
3. Separate remaining data into short and long responses.
   a. Short responses are identified by research team as maximum character length of comments in which there is strong likelihood that only one idea is conveyed.
   i. Further divide short responses into those that are contextually significant and those that are conceptually significant.
   ii. Identify idea clusters, key words aligned with each idea cluster, and automate coding using key words.

Table 1. Key Terms and Definitions.

| Key Term                        | Definition                                                                 |
|--------------------------------|---------------------------------------------------------------------------|
| Contextual significance         | Appropriateness of comment to question asked                              |
| Conceptual significance         | Appropriateness of comment to topic of inquiry                             |
| Codeable universe               | Total amount of qualitative data appropriate for coding                   |
| Test environment                | A purposefully selected sample of the data set used to test macro parameters appropriate to analysis needs |
| Nonmeaningful comment           | Entries of nonsensical keystrokes, for example, @$# or . . . /            |
| Meaningful but not useful comment | Entries that are meaningful but do not offer content of analytic significance, for example, “no comment” |
| Macro                           | Mini programs, often using Visual Basic, that automate predictable, repetitive sets of keystrokes or functions |
| LEN                             | A formula used within Excel to calculate number of characters in a cell    |
| Simple data                     | Free text entries short in length, conveying a single descriptive idea, for example, “today is sunny” |
| Compound data                   | Free text entries that combine two or more simple descriptive ideas, for example, “today is sunny but tomorrow it will rain” |
| Complex data                    | Free text entries that read more like paragraphs and convey concepts not easily separated by key words or simple ideas |
| Analytic unit                   | The smallest unit of analysis in your data set                             |
| Idea cluster                    | Key words grounded in the data and analytically related. In interview transcripts or traditional text-based data sets, idea clusters appear as code families |
| Structural data characteristics | Characteristics used to differentiate data groups based on such factors as length of free text comment (number of characters) and type of significance (contextual or conceptual) |
| Internal data characteristics   | Characteristics used to differentiate content level of a comment, whether simple (short, easily identified single idea), compound (two short, easily identified ideas in a single comment), and complex (nuanced comment not easily assigned to a code using template-based techniques) |
b. Long responses are identified by research team as comments of character length most likely to allow for compound or complex responses.
   i. Further divide long comments into simple, compound, or complex responses. Each of these will have subdivisions for comments that are either contextually or conceptually significant.
   ii. For both simple and complex responses, identify idea clusters, key words aligned with each idea cluster, and automate coding using key words.
   iii. For complex responses, begin your qualitative codebook with codes appropriate to known idea clusters. Continue to build the codebook using an emergent process by which new idea clusters arise from reading and reviewing the data.

4. Analyze data coding patterns for significance.
5. Compare and contrast findings of qualitative significance with findings of quantitative significance to more fully understanding findings and to explore apparent inconsistencies among findings.

Results

The RSM approach was developed through three cycles of problem-solving based on core challenges endemic to large data sets of survey-based free text comments. The results below are organized around these three challenges. For each challenge, we briefly outline the solutions developed and then demonstrate what application of our solution looks like when adapted through application to a specific data set (see Table 2).

Challenge 1: Reduce Noise—Eliminate or Isolate Data That Are Not Analytically Useful

Solution. Large data sets of free text comments often contain nonuseful information (Cunningham & Wells, 2017). If respondents perceive data entry is required, the likelihood of nonuseful information increases. That information, or “noise,” can be distracting, time-consuming to manage, and disruptive to analysis. To answer this challenge, we created a process able to accomplish two tasks: elimination of nonmeaningful entries—noise—and isolation of comments that are meaningful but not analytically useful.

Data includes noise, that is, nonmeaningful or meaningful but not useful entries

1. Organize data as simple, tabular set.
2. Use macro to isolate nonmeaningful data based on character length.
3. Use macro to isolate meaningful but not useful data, based on character length and key words.

Data are of variable type, length, and complexity

1. Group data based on type of comment.
2. Within groups, separate data into subgroups of simple, compound, and complex data based on character length.
3. Use macro on subset of simple data to assist with identifying key words useful for automated coding.
4. Data less than 6 characters in length were removed.

Data are of variable type, length, and complexity

1. Group data based on type of comment.
2. Within groups, separate data into subgroups of simple, compound, and complex data based on character length.
3. Use macro on subset of simple data to assist with identifying key words useful for automated coding.
4. Data less than 6 characters in length were removed.

With great variability, it is difficult to develop a coherent coding strategy

1. Identify simplest analytic unit based on data and purpose of survey.
2. Within each subgroup, use key words to identify suitable idea clusters.
3. Automate coding among simple and compound subgroups informed by idea clusters.
4. Use an emergent process to hand code complex data and any data left uncoded through the automated process.

Results

The RSM approach was developed through three cycles of problem-solving based on core challenges endemic to large data sets of survey-based free text comments. The results below are organized around these three challenges. For each challenge, we briefly outline the solutions developed and then demonstrate what application of our solution looks like when adapted through application to a specific data set (see Table 2).
will for most. In application, researchers are encouraged to consider the most likely useful responses to their open-ended questions when establishing the appropriate length for a noise reducing macro.

After eliminating noise, we found two common types of “meaningful but not useful” entries. The first type involved comments that conveyed the respondent had nothing more to offer, such as “n/a” or “no comment.” The second type involved comments that mirrored the associated scaled numeric question and offered no additional information, such as following a rating of 5, on a scale of 1–5, followed by the free text comment of “excellent.” We found there was a clear subset of free text comments that were duplicative of quantitative ratings making them not useful in the sense of being able to add information to our analyses. Not all survey designs will generate this kind of meaningful but not useful data. Deciding whether a comment is meaningful but not useful will therefore depend on the research questions being answered and the particular survey instrument.

Removing meaningful but not useful comments required identifying a set of key words and then creating macros able to isolate data that met both requirements of being under a specified character length and containing one or more of our key words. Trial and error supported the creation of macro LEN15—isolating comments of fewer than 16 characters and incorporating key words the research team identified as non-responsive (e.g., no comment) or ratings equivalents (e.g., excellent). To identify key words, we first sorted the data in our test environment by entry length and then skimmed it for patterns. While this may seem an informal approach to key word identification, macros created this way correctly identified the data groups of interest more than 97% of the time. Each research team will need to make an independent decision regarding how much time and effort to spend on isolating meaningful but not useful comments prior to coding.

**Application.** Reviewing our sample data set, we found that many nonmeaningful entries were short bursts of random keystrokes. We therefore created macro LEN5. After using LEN5 in our test environment, we applied it to the full codeable universe. Our intention to remove these entries from our data set led to the conservative decision to physically review all removed data. Hand coding of the data isolated using LEN5 showed that among the 67,207 comments removed using this process, fewer than 1,865 were meaningful. The meaningful comments removed were 2.7% of the total number of comments removed and 0.03% of the codeable universe. In both cases, these met our team identified standard of acceptable since they were below the 5% threshold.

After eliminating noise, we next sought to eliminate meaningful but not analytically useful comments: Those that were nonresponsive (e.g., “no comment”) and those that were rating equivalents (e.g., “excellent”). These were removed using LEN15, as described above. However, through this process, we identified a third type of comment appropriately isolated and removed: Comments that were meaningful but of limited use. SAM feedback surveys were designed to understanding learner experiences taking the SAM and their assessment of the quality of the SAM. While examining meaningful but not useful data, coders identified a series of words and phrases that indicated a new subset. There were a significant number of comments that addressed no question in particular but offered a general impression of positive or negative SAM experience. Examples of such comments included “this was really helpful,” “this was a waste of my time,” and “I plan to use this in my practice.” We therefore created two additional macros: LEN50 pos and LEN50 neg. The greater comment length was necessary to allow for expression of a complete thought. Through application in our test environment, we found comments of fewer than 51 characters that included one or more of our key words related to generally positive or negative SAM experience were unlikely to contain additional, analytically useful information. However, isolating these comments and being able to report on the relative frequency of positive comments to negative comments (2 to 1) was of use to survey designers.

Removal of blank cells in conjunction with the use of these four macros—LEN5, LEN15, LEN50 pos, and LEN50 neg—allowed us to significantly reduce the size of our data set requiring analysis. Our codeable universe shrank from a potential 5.2 million comments to a more manageable codeable universe of 440,579 free text comments (see Table 3).

### Table 3: Reduction in Size of Data Set Resulting From Use of Macros.

| Name of SAM                  | Number of Response Entries, Meaningful and Nonmeaningful | Number of Meaningful and Useful Comments | Percent Reduction in Data Set (%) |
|------------------------------|----------------------------------------------------------|-----------------------------------------|----------------------------------|
| Asthma                       | 45,081                                                   | 37,026                                  | 18                               |
| Coronary artery disease      | 34,600                                                   | 27,038                                  | 22                               |
| Care of vulnerable elderly   | 56,019                                                   | 35,948                                  | 36                               |
| Cerebrovascular disease      | 21,653                                                   | 15,586                                  | 28                               |
| Childhood illness            | 56,670                                                   | 34,938                                  | 38                               |
| Depression                   | 28,811                                                   | 23,172                                  | 20                               |
| Diabetes                     | 60,817                                                   | 49,091                                  | 19                               |
| Health behavior              | 39,184                                                   | 26,530                                  | 32                               |
| Heart failure                | 18,914                                                   | 14,589                                  | 23                               |
| Hospital medicine            | 1,715                                                    | 1,187                                   | 31                               |
| Hypertension                 | 52,239                                                   | 40,789                                  | 22                               |
| Maternity care               | 29,950                                                   | 19,840                                  | 34                               |
| Mental healthcare            | 3,545                                                    | 2,588                                   | 27                               |
| Pain management              | 69,662                                                   | 45,285                                  | 35                               |
| Preventive care              | 30,969                                                   | 22,624                                  | 27                               |
| Well child care              | 66,281                                                   | 44,348                                  | 33                               |
| Total                        | 616,110                                                  | 440,579                                 | 28                               |

Note. SAM = self-assessment module.

**Challenge 2: Addressing Variability by Structuring Based on Type and Complexity**

**Solution.** The potential cacophony of large amounts of free text comments can be structured in ways that complement the natural variability within the data. Structuring allows subtle
transformations of the data that ease analysis without disrupting the internal significance, relationships, or character of the data set. Consider the types of questions asked within your survey, the length of comments offered, and the responsiveness of those comments. Questions that are more directive are more likely to elicit responses that are contextually significant. Comments in relation to these questions should be treated as unique subsets of data. Questions that encourage conceptually significant responses are less prescriptive, eliciting responses to ideas within the survey that are not limited to the particular question asked. They form a second kind of data subset.

In addition to variation in response type (contextual or conceptual), comments vary in length. Structuring data based on length is useful in two ways. First, using a macro to isolate comments only a few words in length can help to identify key words that are used repeatedly and likely important to ideas conveyed consistently throughout your data set. Within our sample data, we found comments of fewer than 21 characters in length (LEN20) were most useful for this purpose. While the length of comments useful for this purpose will vary study to study, we recommend less than 21 characters in length as a starting point. Once key words have been identified, the team can generate a second macro, longer in length, able to allow for expression of complete thoughts yet still relying on the use of key words. This second macro is then defined by two parameters: inclusion of key words and restriction of character length.

The character length used to support the second macro will depend on your data set and analysis needs. Our team identified simple data as comments with length fewer than 51 characters (LEN50). Within our sample, these responses typically offered one point of analytic significance, easily identified. Comments in our sample data set of over 50 characters in length were either compound or complex. Compound comments appear as two simple comments in one. They represent easily accessed ideas and follow a predictable pattern as a direct response to a survey question (contextual significance) or to a survey domain (conceptual significance). Complex comments are less predictable or formulaic and require hand coding. They tend to be comments about the presence or absence of a domain within the survey, offering a general world view, or comment on the survey activity. They can be nonresponsive to the question asked and yet analytically significant, requiring attention. Within our data set, comments of character length greater than 150 provided the most opportunity to develop complex responses. As previously stated, while character length is likely dependent on survey questions and respondents, LEN50 and LEN150 are strong starting points for your data exploration.

Application. The SAM feedback surveys used to create our sample data set employed six types of questions generating data relative to SAM content (e.g., content was presented at appropriate level), the skill tested (e.g., this activity helped me to improve my skills in diagnosing), views regarding learning objective, views regarding learning modality, comments specific to the SAM user interface (e.g., the program handles “natural language” input appropriately), and overall opinions regarding the SAM as a whole (e.g., I have a favorable impression of this exercise).

Questions that were more directive, such as yes/no questions (e.g., do you feel there was any bias toward a particular product or service in this activity), or questions that asked variations of “what one to three things did you learn from this activity” were more likely to result in comments classified as contextually significant simple data (LEN50) and able to be treated as unique subsets of our data set. Longer responses to these questions were most often compound (LEN150) yielding predictable answers to a specific question. Comments structured as informal or responding to technological difficulties in the SAM exercise were conceptually significant. Simple and compound responses that were conceptually significant were combined into a single subset. In these cases, we used the strategy for identifying key words, described in Challenge 1. Simple and compound responses coded using key words grouped in meaningful clusters can be understood alone or in combination with the numeric responses from the paired scaled survey question. Complex responses were isolated for hand coding.

**Challenge 3: Developing a Coherent Coding Strategy That Makes Use of Automation**

**Solution.** Developing a coherent coding strategy first requires that members of a research team agree on how they define an analytic unit. Simple data, mentioned above, contain one analytic unit or idea. Compound data contain two or more analytic units that are easily discerned from each other. Both of these data types are easily coded using an automated process based on idea clusters — groups of key words that all related to a single idea or concept asked about in your survey and/or present among free text comments. How a research team defines an idea cluster and how broadly the team operationalizes the idea cluster will depend on the size of the data set (Namey et al., 2008). Analytic units within complex data are too interactive or overlapping to allow for easy and clean separation. Automated coding of complex data using this process is not recommended.

Among each subset of simple data, a hand review of data identified using LEN20 should enable easy identification of key words related to survey idea clusters. We found the best application of this approach to be iterative. First sorting data using LEN20, then identifying potential key words, then temporarily isolating data identifiable through those key words to determine whether there are additional patterns visible in the remaining data. Temporarily isolating data related to larger patterns allow for easier recognition of potentially useful but smaller data patterns. This should be done until saturation is reached, that is, that point at which you are continuing to review the data but no new key words, or idea clusters, are being identified. This will typically happen on review of 1–10% of your data, dependent on the size of your subset. Key words identified using this process can then be used to auto-code both simple and compound data with the help of any number of programs. We discuss our use of the free
OpenRefine software, Version is 2.5 below (http://openrefine.org); however, use of advanced Visual Basic within a tabular worksheet environment can accomplish the same results.

Using OpenRefine, macros with Microsoft Excel, or other software options, simple frequencies related to the appearance of key words among data subsets will allow you to identify the prevalence of idea clusters within your data as well as the potential discovery of smaller nested clusters. In addition, this level of analysis and the transmutation of your free text comments into prevalence of idea clusters facilitates integration of qualitative and quantitative findings. When our team used this technique to analyze our data set, we found that after two iterations, our process successfully coded over 95% of the analytically significant simple and compound data.

Key words identified using this process can also inform the hand-coding process for any isolated subsets of complex data. In this case, the key words become codes and idea clusters are code families. Together, they form an a priori codebook for template-based coding (Crabtree & Miller, 1999). While this may be sufficient, some research teams will also want to employ a grounded theory approach, leaving their codebook open to codes that naturally emerge as significant during the process of data review (Crabtree & Miller, 1999). An emergent process is likely to enable a deeper understanding of the ideas present within your data and may point to domains of significance not represented in simple and compound comments.

**Application.** Our team identified key words within simple and complex data subsets using OpenRefine software. Freely available online (openrefine.org), this software allows easy manipulations of your data using an Internet–browser interface to explore data while it remains on your local drive. This can also be accomplished through visual review. Use of software reduces iterations necessary to reach saturation, but it is not necessary for applying this method. In most cases, the analytic output possible using this method in conjunction with a tabular environment and use of macros will be sufficient.

Those research teams interested in more sophisticated analyses, such as the grouping of similarly structured free text comments (e.g., all comments reflecting the syntax “I found it most useful when . . .”), would find software programs such as OpenRefine advantageous. OpenRefine allows easy identification of simple word patterns within your data. It will group like words or word patterns and ask the user to identify whether this group is a true group. It can do so with entries that are visually similar (e.g., written plan, Written plan, and written plans) or entries that are phonetically similar (e.g., counseling, counseling, and counselling), and organizes potential groups by prevalence, allowing the user to determine at what point the prevalence is so low as not to be useful.

After identifying what we felt to be a complete set of key words, OpenRefine then permitted the use of General Refine Expression Language (GREL) expressions to determine the prevalence of individual key words within our data set as well as idea clusters. GREL allows the user to create a single script, similar to many macros in one, that can identify both the frequency of key words among responses and the frequency of a key word group (idea cluster) among responses. The same process can be accomplished in a tabular software environment. Our team found use of OpenRefine required less skill with Visual Basic, fewer steps, and an easier user interface.

Calculating frequency of appearance by response prevented us from potential misrepresentation of the saturation of idea clusters within our data. For example, within the Asthma SAM, one learning objective was to understand the need for “written action plans.” Imagine there are 10 responses in your data set and three key words within an idea cluster of interest: written, action, and plan. If “written” appeared 5 times, “action” appeared 5 times, and “plan” appeared 7 times, you might accidentally add these frequencies and find that the idea cluster appeared 17 times within your data—not possible among 10 responses. However, with focused attention to the frequency of an idea cluster, and independent key words, you might find three occurrences of “written action plan,” two occurrences of “action plan,” and two occurrences of “written plan,” correctly identifying the idea cluster as present in 70% of your data. While these analyses are possible within simple tabular environments, the ability of OpenRefine to simplify the commands and steps necessary to achieve this outcome was found advantageous by our team.

Using idea clusters and GREL expressions within OpenRefine allowed us to automate coding of 95% of our data. Roughly, 30,800 free text comments required hand coding for which we used both a template-driven and emergent process. Results from this analysis have been published elsewhere (Brooks et al., 2017).

**Discussion**

Using the RSM approach outlined above, within a week, our research team of three was able to reduce a potential 5.2 million free text comments to a codeable universe of 616,000 comments (440,500+ meaningful and useful, 175,400+ meaningful and of limited use). Of these, only 30,800 required hand coding, that is, approximately 5% of our meaningful and useful free text comments. This remainder, equivalent to approximately 1,000 pages of text, was hand coded by our team in less than 160 hr. RSM of large amounts of free text comments is feasible and can yield actionable results. The easy to follow steps that we have outlined above can help researchers to divide data into subsets, most of which are amenable to carefully planned automated coding, informed by a quick review of a small portion of data. As described above, analyses of survey findings can then be enhanced when quantitative data are viewed in relation to the frequency of idea clusters within and across data subsets as well as patterns found within or across idea clusters. Findings not captured or understood through numeric answers alone can be exposed through variations in cluster frequencies, number of clusters per subset, and discovered idea clusters bearing little or no correlation to survey-relevant domains.
Limitations

Analysis of survey results is always open to respondent bias. Our data set is no different. However, we were unable to adjust our process for potential respondent bias. Our data set was collected over a 10-year period during which the fielding of the survey was adjusted. Some years, responding to survey questions was voluntary, and some years, it was mandatory. There also appeared to be times when respondents were unclear if free text comments were mandatory. While we might expect a negative survey bias, our analysis of generally positive and generally negative comments showed an overall positive opinion of SAMs at a rate of 2 to 1 (Brooks et al., 2017).

A second bias that might appear within the data is the over representation of particular respondent views. With up to 20 opportunities to submit free text comments, it is possible that one person might submit 20 comments while another might submit 1. Tracking unique identifiers related to each comment can control for this potential issue.

The surveys that produced our data were both varied and focused. The scaled questions asked were relatively limited in focus (e.g., content was presented at appropriate level, and this activity helped me to improve my skills in diagnosing). Therefore, the free text opportunities associated with these questions were more likely to be limited in scope. In addition, the homogeneity of our sample, all board certified family physicians, created a certain predictability to viewpoints offered that made key word identification easy. Surveys with a broader focus or greater open-ended nature may not be well matched with the RSM method.

Conclusion

Free text comments in surveys are able to provide a rich source of data that complements numerically based data capture. The ability to efficiently incorporate free text comments into survey analyses can prevent the loss of potentially important findings and may open the door to researchers previously weary of collecting large mixed data sets. This is particularly true for surveys that are structured or semistructured, thereby limiting the potential breadth of ideas shared through free text comments. Using this approach to analyze survey-based free text comments will not only ease the time and resource burden often cited for not integrating this type of data within final analyses, it can also assist in the integrated reporting of findings or the design of free text questions to confine answers to contextual and simple or compound responses. The RSM approach can make analyzing and integrating qualitative data obtained through the inclusion of open-ended questions realistic for health services researchers. It creates the potential to enhance and deepen our understanding of many issues studied via survey-based research tools, including patient satisfaction, physician experience, and workforce issues.

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.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Rebecca S. Etz http://orcid.org/0000-0001-6906-7738

References

Bankauskaite, V., & Osmo, S. (2003). Why are people dissatisfied with medical care services in Lithuania? A qualitative study using responses to open-ended questions. *International Journal for Quality in Health Care, 15*, 23–29.

Bracher, M., Corner, D. J., & Wagland, R. (2016). Exploring experiences of cancer care in Wales: A thematic analysis of free-text responses to the 2013 Wales Cancer Patient Experience Survey (WCPES). *BMJ Open*, 6, e011830. doi:10.1136/bmjopen-2016-011830

Bracher, M., Wagland, R., & Corner, J. (2014). Exploration and analysis of free-text comments from the 2013 Wales Cancer Patient Experience Survey (WCPES). Southampton, GB: University of Southampton. Retrieved November 14, 2017, from https://eprints.soton.ac.uk/372729/1/WCPES%2520Report%2520FINAL.pdf

Brooks, E. M., Gonzalez, M., Eden, A. R., O’Neal, J., Sabo, R. T., & Etz, R. S. (2017). What family physicians really think of maintenance of certification Part II activities. *The Journal of Continuing Education in Health Professions, 37*, 223–229.

Crabtree, B. F., & Miller, W. L. (Eds.). (1999). *Doing qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Cunningham, M., & Wells, M. (2017). Qualitative analysis of 6961 free-text comments from the first National Cancer Patient Experience Survey in Scotland. *BMJ Open*, 7, e015726. doi:10.1136/bmjopen-2016-015726

Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods, 5*, 307–336. doi: 10.1177/109442802237114

MacRae, J., Darlow, B., McBay, L., Jones, O., Stubbe, M., Turner, N., & Dowell, A. (2015). Accessing primary care Big Data: The development of a software algorithm to explore the rich content of consultation records. *BMJ Open*, 5, e008160. doi:10.1136/bmjopen-2015-008160

Maramba, I. D., Davey, A., Elliott, M. N., Roberts, M., Roland, M., Brown, F., . . . Campbell, J. (2015). Web-based textual analysis of free-text patient experience comments from a survey in primary care. *JMIR Medical Informatics, 3*, e20. doi:10.2196/medinform.3783

Marcinowicz, L., Chlabicz, S., & Grebowksi, R. (2007). Open-ended questions in surveys of patients’ satisfaction with family doctors. *Journal of Health Services Research and Policy, 12*, 86–89.

Moynihan, R., Nickel, B., Hersch, J., Doust, J., Barratt, A., Beller, E., & McCaffery, K. (2015). What do you think overdiagnosis means? A qualitative analysis of responses from a national community survey of Australians. *BMJ Open*, 5, e007436. doi:10.1136/bmjopen-2014-007436

Namey, E., Guest, G., Thairu, L., & Johnson, L. (2008). Data reduction techniques for large qualitative data sets. In G. Guest & K.
MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137–161). Lanham, MD: Altamira.

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In J. Gennari & B. Porter (Eds.), Proceedings of the 2nd international conference on knowledge capture (pp. 70–77). New York, NY: ACM.

O’Cathain, A., & Thomas, K. J. (2004). “Any other comments?” Open questions on questionnaires—a bane or a bonus to research? *BMC Medical Research Methodology, 4*, 25.

Rich, L. R., Chojenta, C., & Loxton, D. (2013). Quality, rigour and usefulness of free-text comments collected by a large population based longitudinal study—ALSWH. *PLoS One, 8*, e68832.

Richards, S. H., Campbell, J. L., Walshaw, E., Dickens, A., & Greco, M. (2009). A multi-method analysis of free-text comments from the UK General Medical Council Colleague Questionnaires. *Medical Education, 43*, 757–766.

Riiskjaer, E., Ammentorp, J., & Kofoed, P. (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.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*, 1064–1082. doi:10.1111/ajps.12103

Wiedemann, G. (2013). Opening up to big data: Computer-assisted analysis of textual data in social sciences. *Historical Social Research/Historische Sozialforschung, 38*, 332–357. Retrieved from JSTOR: http://www.jstor.org/stable/24142701