$\textcircled{S}$SAGE

How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes

Greg Guest[1], Emily Namey[1], and Kevin McKenna[1,*]

Abstract
Few empirical studies exist to guide researchers in determining the number of focus groups necessary for a research study. The analyses described here provide foundational evidence to help researchers in this regard. We conducted a thematic analysis of 40 focus groups on health-seeking behaviors of African American men in Durham, North Carolina. Our analyses revealed that more than 80% of all themes were discoverable within two to three focus groups, and 90% were discoverable within three to six focus groups. Three focus groups were also enough to identify all of the most prevalent themes within the data set. These empirically based findings suggest focus group sample sizes that differ from many of the "rule of thumb" recommendations in the existing literature. We discuss the relative generalizability of our findings to other study contexts, and we highlight some methodological questions about adequate sample sizes for focus group research.

[1] FHI 360, Durham, NC, USA
[*] Kevin McKenna's current affiliation is Duke Clinical Research Institute, Durham, NC, USA

Corresponding Author:
Greg Guest, FHI 360, 359 Blackwell Street, Suite 200, Durham, NC 27701, USA.
Email: gguest@fhi360.org

Introduction

Focus groups are commonly used across a wide range of research disciplines, including health sciences, marketing, communications, and virtually all fields of behavioral and social sciences. Despite this widespread use, little attention has been given to methodological aspects of the technique, beyond the formation and conduct of focus groups (Agar and MacDonald 1995; Carey 1995; Duggleby 2005; Kaplowitz and Hoehn 2001). One of the largest gaps in the focus group literature concerns the question of sample size: How many focus groups are needed in a study to adequately address a research objective?

Seventy years after the method was introduced, researchers must still rely on rules of thumb and personal judgment when deciding how many focus groups to include in a study. Similarly, reviewers and funders have little basis for determining whether a study includes an adequate number of focus groups. The analyses we present in this article begin to address the need for empirically based recommendations on focus group sample size.

Focus Group Sample Size Recommendations

We reviewed textbooks that cover general qualitative research or focus group methodology (bibliography available in Online Appendix). Of the 62 books reviewed, 42 provided no guidance on the number of focus groups needed for a study (many did not even cover the topic of focus groups), six recommended saturation as the key determinant, 10 provided some form of numeric recommendation, and four mentioned saturation and included a number. Focus group sample size recommendations from these books and a number of journal articles range from as few as two focus groups per study to more than 40 (Fern 1982; Greenbaum 1997; Kitzinger and Barbour 2001; Krueger and Casey 2015; Morgan 1996; Powell and Single 1996; Vaughn et al. 1996). One commonly cited guideline is that focus group research requires at least two groups for each defining demographic characteristic (Barbour 2007; Carey 1995; Knodel 1993; Krueger and Casey 2015; Morgan 1988; Ulin et al. 2005). None of these recommendations are supported by empirical data.

Nonprobability Sampling and Saturation

Many published works cite "theoretical saturation" as the primary method for determining nonprobability sample sizes in qualitative research (e.g.,

Bluff 1997; Byrne 2001; Fossey et al. 2002; Guest et al. 2006; Morse 1995; Sandelowski 1995). Theoretical saturation was first defined by Glaser and Strauss (1967:61) as the point at which "no additional data are being found whereby the [researcher] can develop properties of the category." Glaser and Strauss's definition, however, is intended for a grounded theory approach, in which theoretical models are developed and constantly compared against incoming data. For grounded theorists, theoretical saturation refers to the point at which the theoretical model being developed stabilizes. Because not all qualitative data analysis employs a grounded theory approach, Guest and colleagues (2006:65) and others use the broader term "data saturation," which is defined as "the point in data collection and analysis when new information produces little or no change to the codebook."

The concept of saturation has become the gold standard by which sample sizes for qualitative inquiry, including focus groups, are determined (Guest et al. 2006; Guest and MacQueen 2008). However, using the concept as a practical reference for estimating sample sizes is problematic because most research requires such estimation *before* a study is implemented (Charmaz 2014; Cheek 2000). By definition, saturation can be determined only during or after data analysis.

Some scholars have addressed this issue using mathematical modeling or statistical methods. Romney et al. (1986) used factor analysis and mathematical modeling (cultural consensus theory) to show that the confidence level associated with obtaining accurate information is a positive function of the degree of consensus about a particular topic and the number of interviewees. They calculated that as few as four individuals can render accurate information with a high confidence level (.999) if they possess a high degree of knowledge with respect to the domain of inquiry. Galvin (2015) reviewed and statistically analyzed—using binomial logic—54 qualitative studies and identified a similar trend. Like the work of Romney and colleagues, Galvin's analysis illustrates that the probability of identifying a theme is a function of the breadth of its presence within a study population and the number of individuals in a sample. Based on his analysis, the probability of identifying a concept (theme) among a sample of six individuals is greater than 99% if that concept is shared among 55% of the larger study population.

Employing this same logic, Fugard and Potts (2015) developed a quantitative tool to estimate sample sizes for thematic analyses of qualitative data. Their calculation incorporates (1) the estimated prevalence of a theme within the population; (2) the number of desired instances of that theme;
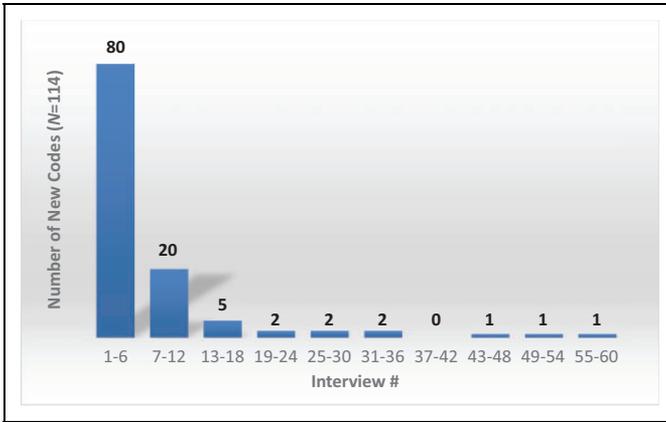
Figure 1. Code creation during analysis of 60 individual interviews (adapted from Guest et al. 2006).

and (3) the desired power for a study. Their tool estimates, for example, that to have 80% power to detect two instances of a theme with a 10% prevalence in a population, 29 participants would be required.

These theoretical models and tools provide general guidance for determining sample sizes for qualitative studies. As with any theoretical model, however, they are limited by their inherent assumptions and absence of empirical data. One of the first studies to bridge this gap, and systematically examine saturation using raw data, was conducted by Morgan et al. (2002). In individual interviews about environmental risks, Morgan and colleagues found that the first five to six interviews produced the majority of new data, and little new information was gained as the sample size approached 20 interviews. Across all four of their interview data sets, approximately 80–92% of concepts were identified within the first 10 interviews.[1] This finding is consistent with the results of a stepwise inductive thematic analysis of 60 in-depth interviews (10 successive batches of six interviews) among female sex workers in West Africa (Guest et al. 2006). Of the 114 themes identified in the entire data set, 80 (70%) turned up in the first six interviews and 100 themes (92%) were identified within the first 12 interviews (Figure 1).

Results from two empirical studies reported by Francis et al. (2010) are congruent with earlier findings (Guest et al. 2006; Morgan et al. 2002). In one study, Francis and colleagues interviewed 14 doctors to identify their beliefs about managing patients with upper respiratory

tract infections. In another study, they interviewed 17 blood relatives of people with Paget's disease of the bone to explore the acceptability of a potential genetic screening service. For both studies, the researchers operationalized saturation as the point, after conducting 10 interviews, when three additional interviews yielded no new themes. Based on this definition, the authors did not reach saturation within the first study among 14 doctors but did reach saturation after 17 interviews in the second study. Due to the small sample sizes, estimating the proportion of themes identified at distinct points in their analysis is difficult. However, in both of the studies, the vast majority of themes were identified within five to six interviews—the point at which the curve begins to sharply flatten out.

Research by Hagaman and Wutich (2017) suggests that cross-cultural research contexts may require more interviews to reach saturation than the aforementioned studies. These authors conducted a study exploring water norms and knowledge across four sites in four countries. They found that fewer than 16 interviews were enough to identify common themes from each of the individual sites, but that 20–40 interviews were needed to identify meta-themes that cut across all sites.

We found only two studies that empirically assessed saturation within focus groups. Coenen et al. (2012) compared two approaches for analyzing focus groups and individual interviews: an inductive (open coding) approach in which themes and codes were generated from the data and a deductive approach in which existing codes—from the International Classification of Functioning, Disability and Health (ICF)—were applied to the data. The authors operationalized saturation as the point during data collection and analysis at which linking the concepts of two consecutive focus groups or individual interviews revealed no additional second-level categories. Saturation was reached after performing five focus groups and eight individual interviews using the inductive approach and five focus groups and eight individual interviews using the deductive approach. In a study intended to validate the ICF—and using the same method of assessing saturation as Coenen et al. (2012)—Kirchberger et al. (2009) found that saturation was reached after conducting eight focus groups. A summary of these two data-driven studies, as well as those based on individual interviews, is presented in Table 1.

Our study builds on the existing literature by focusing explicitly on inductive thematic analysis within a study specifically designed to assess saturation. The saturation-focused objectives of the research allowed us to justify a large sample size and keep the procedures and

Table 1. Summary of Saturation Findings from Empirical Studies.

Study authors	Saturation definition	Findings
Individual interviews		
Morgan et al. (2002)	Not defined	Five to six interviews for most concepts
		In all four sets of interviews, approximately 80–92% of concepts identified within 10 interviews
Guest et al. (2006)	The proportion of identified themes at a given point in analysis divided by the total number of themes identified in that analysis	Six interviews to reach 70% saturation
		12 interviews to reach 92% saturation
Francis et al. (2010)	The point, after conducting 10 interviews, when three additional interviews yield no new themes	Most themes in both studies identified within five to six interviews
		Saturation reached within 17 interviews in one study and not reached in 14 interviews in a second study
Coenen et al. (2012)	The point at which linking concepts from two consecutive focus groups or individual interviews reveals no additional second-level categories	Inductive approach: 13 interviews to reach saturation
		Deductive approach: 8 interviews to reach saturation
Hagaman and Wutich (2017)	The number of interviews required to identify the most common themes in a total of three interviews	<16 interviews at site level
		20–40 interviews to identify cross-cultural metathemes
Focus groups		
Kirchberger et al. (2009)	As per Coenen et al. (2012)	Deductive approach: eight focus groups to reach saturation
Coenen et al. (2012)	As above	Inductive approach: five focus groups to reach saturation
		Deductive approach: five focus groups to reach saturation

context surrounding data collection and analysis consistent to enhance methodological rigor.

Methods

This research is part of a larger study funded by the Patient Centered Outcomes Research Institute, which supports methodological investigations on engaging patients in health and health-care research (see www.pcori.org). We conducted two prestudy focus group discussions among members of the African American community in Durham, North Carolina, to solicit their opinions on the health issues most in need of research within their community. We chose health-seeking behavior among African American men as the topic of the focus group discussions (substantive findings from the study presented elsewhere), based on the funder's objectives and the community's feedback. The study was reviewed and approved by FHI 360's Protection of Human Subjects Committee. Oral informed consent was obtained from each participant before initiation of data collection. Each study participant was provided an incentive of US$40.

Data Collection

Eligibility criteria included being a (self-identified) African American/ black man, aged 25–64, and a resident of Durham, North Carolina. The study team recruited participants through Craigslist, through flyers posted in public areas, and through peer recruitment. For each focus group, we attempted to include eight individuals—the modal recommendation for group size in the literature—which also kept the focus group size consistent across each data collection event. We overrecruited for each focus group to offset any no-shows.

A draft of the focus group instrument was pretested among five men from the target population and revised based on their responses. The final instrument contained 13 open-ended questions. Between January and May 2013, we conducted 40 focus groups—a number within the upper range of recommendations in the literature, enabling us to reach the data saturation point with confidence. One researcher facilitated all of the discussions. She followed the instrument structure consistently and probed responses to questions, but she did not introduce any information learned in previous focus groups as one typically would in inductive qualitative research. This was done to treat each focus group as a "new" and unique event to facilitate

the methods objectives of the analysis. All focus group discussions were audio recorded.

Focus group instrument (abridged)
1. How long have you lived in Durham, and have you seen a physician in the past year? (warm-up question)
2. What do you think are the most common health problems in the African American community in Durham? (free-listing question)
3. Let's imagine that I have never interacted with the medical system here. Describe for me what a typical doctor visit is like.
4. How do you feel about the typical doctor visit here, as you've described it?
5. Please tell me some positive things about a doctor visit.
6. Please tell me some negative things about a doctor visit.
7. What motivates you to see a doctor?
8. Think back to a time when you could/should have gone to the doctor but didn't. What prevented you from going?
9. Research suggests that men are afraid of some medical procedures, particularly prostate exams.

Data Analysis

The recordings were transcribed verbatim, using a transcription protocol developed for thematic analyses (McLellan et al. 2003). After verifying the accuracy of the transcripts, two members of the study team (including the data collector) created a content-driven, thematic codebook using an iterative process (MacQueen et al. 2008). Both analysts independently reviewed each focus group transcript, identified new themes, and created new codes corresponding to those themes.[2]

All codes were defined and agreed upon using the template outlined by MacQueen et al. (2008). Analysts used NVivo 10 (QSR 2012) to apply codes to the transcripts (coding of text segments was not mutually exclusive). After each transcript was coded, analysts discussed emergent themes and compared code application. All coding discrepancies were resolved through discussion to create a consensus-coded file, and the codebook was revised to add new codes or reflect changes to code definitions. Changes to the codebook were logged after the analysis of each transcript. Analysts were not blinded to the study's purpose, since we had no a priori hypotheses regarding saturation.

We assessed data saturation using a methodology similar to that employed by Guest et al. (2006). We documented the incremental progression of theme identification (i.e., codebook content) by examining the

codebook log to determine when each code was identified. We then documented code frequencies (i.e., the number of focus groups in which a code was identified at least once). We tallied the number of codes created across all 40 transcripts and determined the point when 80% of all thematic codes had been identified. We also analyzed the data based on a 90% metric to provide a more robust measure of saturation.

To investigate the possibility of a temporal bias in our analysis, we randomly ordered the 40 focus groups 10 times[3] using an online random sequence generator. We replicated the saturation analyses described above for each of the 10 randomly ordered data sets, using matrix queries of code application to indicate when new codes appeared, and then averaged saturation measures.

To determine whether the themes identified early in the analysis were the most salient, we used code frequencies as a proxy measure of salience, and grouped the results into terciles—low, medium, and high frequency. We then determined the percentage of high-frequency codes that were identified in the earlier stages of analysis. We did this for the original data set and all 10 randomly ordered data sets.

Results

The focus group discussions lasted, on average, 1 hour and 50 minutes. Each focus group had six to eight individuals (only one group had six participants), with a mean of 7.75 individuals per group.

Sample

Of the 390 individuals who expressed interest in study participation, 49 did not meet the eligibility criteria. Eleven eligible men consented and were scheduled but did not attend the focus group. Twenty men consented, showed up as scheduled, and received an incentive but did not take part because the cap of eight per group had been met. Our final sample included 310 men.

Study Participants

The median age of the men in our sample was 49 years (Table 2). More than 90% of the men had completed high school, 75% were unemployed, and 78% had an annual household income of less than US$20,000. Although 70% of the men had seen a physician within the past year, 61% were uninsured.

Codebook Development and Changes

We developed 94 content-driven codes (refer to the codebook and code frequencies online), and the codebook remained relatively stable during analysis. Only 15 changes were made: 11 code definitions were expanded to be more inclusive (example below), two code names were revised to reflect expanded definitions, and two codes were deleted. All but one of these 15 changes were made early in the analysis between the second and fifth focus groups.

Code name: Experimentation

Brief definition: Experimentation or research on patients as part of health care.

Full definition: Use this code for mention or discussion of past or current experiences or beliefs about experimentation on human patients within the context of health care. *Also includes perceived research or study of patients that is not explicitly called experimentation.* Use for any reference to the Tuskegee trials or for any other kind of actual or perceived experimentation. *May be couched in terms of research, being studied, or used as a guinea pig. Also use for mention of teaching hospitals as "using" patients in this way.*

Revision (in italics). Expanded code to encompass a broader conceptualization of experimentation to include references to research, studies, guinea pigs, and teaching hospitals, whether actual or perceived.

Data Saturation Points

Almost two-thirds of the 94 content codes were generated from the first focus group (Figure 2). Seventy-nine (84%) of the 94 codes were created and applied during the analysis of the first three focus groups. The 90% saturation mark was reached after six focus groups. With a few exceptions, the identification of new themes and the corresponding creation of codes followed a negative curvilinear pattern, with the slope rapidly approaching a plateau by the fourth focus group.

This pattern persisted, regardless of the order in which we analyzed the data. Our analysis of the 10 randomly ordered data sets revealed that the majority of codes—ranging from 76 (81%) to 85 (90%)—were applied within the first three focus groups. When we averaged the number of new codes applied in each focus group across the 10 randomly ordered data sets, the same negative curvilinear pattern that we had observed in the chronologically ordered data set emerged (Figure 3). On average, 80.3 codes (85%) were applied to the first three focus group transcripts and 85.6 codes (91%) to the first four.

Table 2. Participant Characteristics.

Age (N = 309)[a]		
Mean	47	
Median	49	
Range	25–67	
Highest level of education (N = 306)	#	%
Elementary school	3	1.0
Middle school	15	4.9
High school	205	67.0
Associate's degree/tech college	46	15.0
BA/BS	27	8.8
Graduate degree	5	1.6
Other/trade school	5	1.6
Employment status (N = 299)	#	%
Employed	65	21.7
Unemployed	233	77.9
Retired	1	0.3
Annual household income (N = 299)	#	%
<US$20,000	235	78.6
US$20,000–US$40,000	43	14.4
US$40,001–US$60,000	15	5.0
US$60,001–US$80,000	4	1.3
>US$80,000	2	0.7
Insurance status (N = 247)	#	%
Private	22	8.9
Medicaid	48	19.4
Veterans' benefits	19	7.7
More than one type of insurance	7	2.8
Uninsured	151	61.1
Has primary care physician (N = 242)	#	%
Yes	125	51.7
No	117	48.3
Has seen a physician in past 12 months (N = 307)	#	%
Yes	217	70.7
No	90	29.3

[a]Although our total sample population was 310 men, the N differs for each variable because of different response rates.

Combining the chronologically and randomly ordered data sets, the average number of focus groups required to reach 80% saturation was 2.7 (range two to three groups). To reach 90%, the average number of groups required was 4.3, with a range of three to six groups (Table 3).
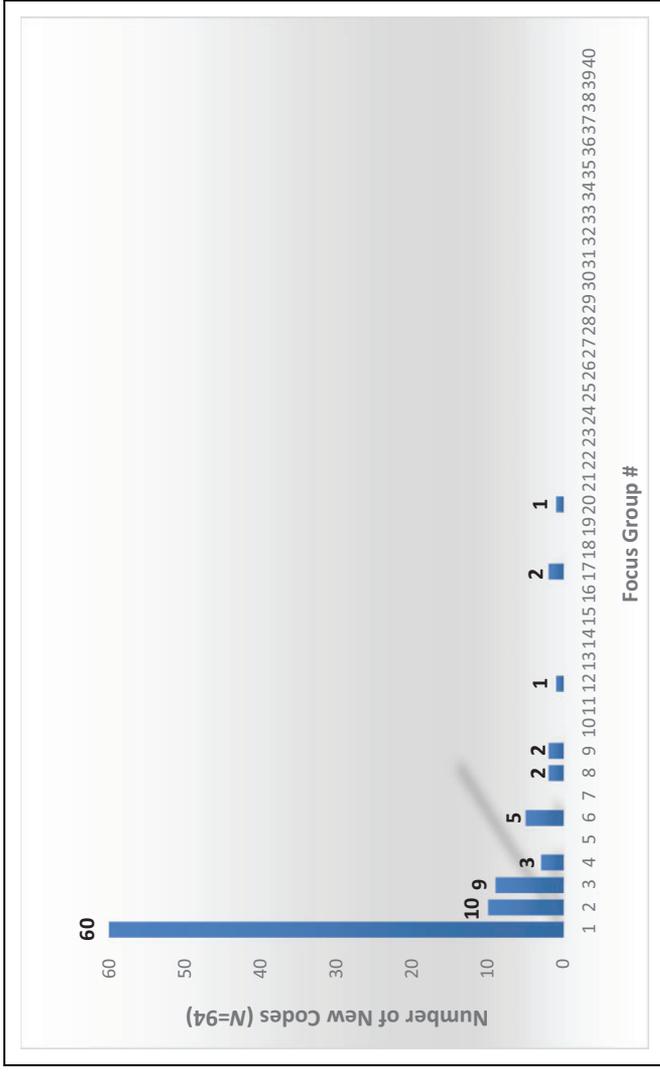
14

Figure 2. Code creation over time, in chronological order of data collection (*N* = 40 focus groups).
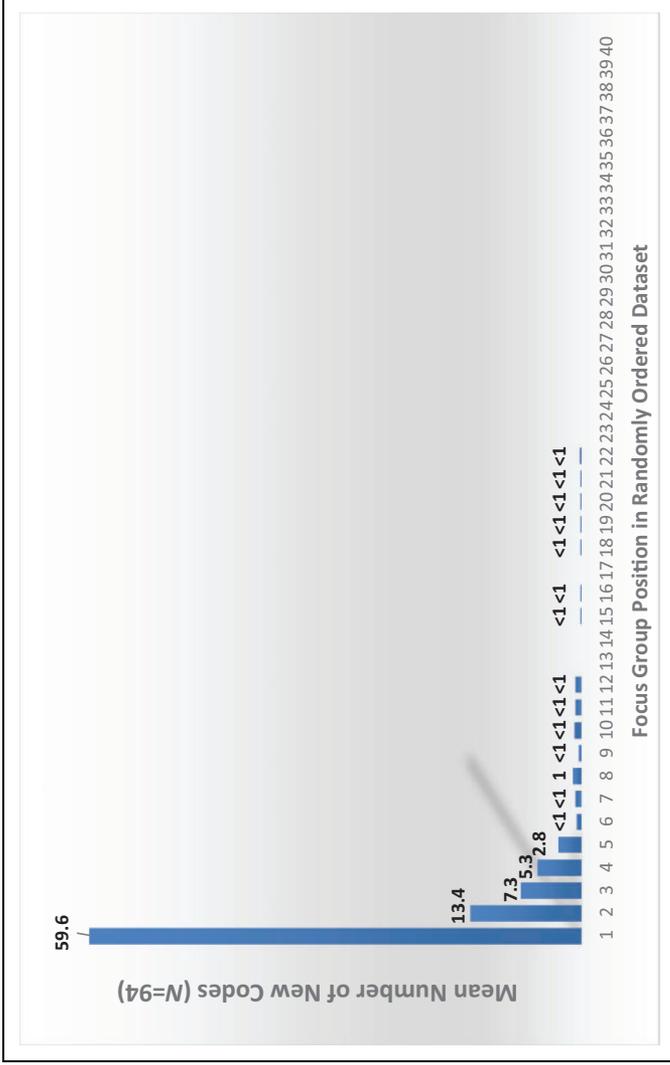
Figure 3. Application of new codes averaged across 10 randomly ordered data sets (N = 40 focus groups).

Table 3. Saturation Points among Data Sets.

Data set	Number of focus groups to reach >80% saturation	Number of focus groups to reach >90% saturation
Chronological order	3	6
Random order #1	3	5[a]
Random order #2	3	5[a]
Random order #3	3	4
Random order #4	3	4
Random order #5	2	4
Random order #6	3	4
Random order #7	3	4
Random order #8	3	4
Random order #9	2	4
Random order #10	2	3
Mean	2.7	4.3

[a]The first four focus groups in these data sets included 89.4% of all codes, just short of the 90% metric.

Code Frequency and Salience

After dividing the 94 codes into terciles—high ($n = 31$), medium ($n = 33$), and low frequency ($n = 30$)—we identified where in the data set the high-frequency codes first appeared. In the original data set, which was sequenced to match the order in which the data were collected and the codes were developed, we found that all of the high- and medium-frequency codes were developed during the analysis of the first three focus groups. The same pattern held in all 10 of the randomly ordered data sets.

Discussion

Our data suggest that a sample size of two to three focus groups will likely capture at least 80% of themes on a topic—including those most broadly shared—in a study with a relatively homogeneous population using a semi-structured guide. As few as three to six focus groups are likely to identify 90% of the themes. To assess the generalizability of these findings, we place them in the context of five general factors that can affect the rate at which saturation is approached: (1) degree of instrument structure; (2) sample homogeneity; (3) complexity of the study topic; (4) study purpose; and (5) analyst categorization style (Guest 2015).

Degree of Instrument Structure

Our study utilized one moderator and a semistructured instrument in which the initial (open-ended) questions were asked verbatim of all groups and in the same order. For qualitative inquiry that doesn't employ a guide or scripted questions, saturation is likely to require more focus groups than in our study. Conversely, more structured elicitation techniques, such as free listing and pile sorting (Weller and Romney 1988), would likely reach saturation sooner.

Sample Homogeneity

Although our sample had some demographic heterogeneity, most participants were similar with respect to education, employment, household income, and health insurance status. More focus groups may be needed to reach saturation as the heterogeneity of the sample population increases.

Topic Complexity

Our study could be classified between simple and moderately complex. We asked the participants about their opinions of and experience with health and the health-care system. On the one hand, the range of health issues and health-care coverage is relatively large (compared to, say, discussing a tangible consumer product), and the subject matter can be technical. On the other hand, this is a simple topic in that participants are relaying their personal experiences and discussing cultural norms. More focus groups may be needed to reach saturation as the complexity of the topic increases or becomes more abstract in nature.

Study Purpose

The purpose of our study was to identify salient themes on a topic. This is a common objective in applied research, so our findings (at least relative to this parameter) should be generalizable to many other focus group studies. Our data may not be generalizable, however, to analyses that seek highly granular themes or aim to determine the maximum variability of participants' responses. Similarly, our findings may not be relevant to research that is more interpretative or doesn't generate quantifiable data such as theme frequencies.

Analyst Categorization Style

Analysts vary in the level of granularity at which they code data—a phenomenon known as the ''lumper–splitter problem'' (Guest et al. 2012). Neither of our two coders had a history of coding at either extreme of the granularity continuum. Their analysis of 40 focus groups generated 94 codes that, based on our experience, are indicative of neither lumping broadly nor splitting narrowly. Studies that use analysts with coding styles at the extremes of the granularity continuum—or who are instructed to code at a very specific or broad level—will find saturation rates to vary accordingly.

In addition to these factors, we might assume that the size of the focus group influences the amount and quality of the data generated. For reasons of consistency, we attempted to limit the size of the groups in our study to eight participants. Although this is the modal size recommended in the literature, focus group sizes often range between six and 12 individuals. This may temper the generalizability of our findings to studies with significantly smaller or larger groups.

Another factor, related to group size, that can potentially affect saturation is the degree of interpersonal dynamics within a group. A few highly disruptive or vocal participants, for example, can reduce the variability of responses within a focus group. This is why good moderating skills are essential to the reliability and validity of focus group research. The influence of group dynamics, however, is likely to average out over time.

Conclusion

Our analyses revealed that, within our data set, more than 80% of all themes were discoverable within two to three focus groups and 90% of themes could be discovered within three to six focus groups. Also, we were able to identify the most prevalent themes within our data set with only three focus groups. These findings build the evidence base about adequate sample sizes for focus group research. At the same time, we recognize that our findings may not apply to all contexts of focus group research. Many questions remain. How does the degree of heterogeneity within a focus group, the complexity of a topic, or the size of a focus group affect the saturation rate and the nature of the data generated? How much does the amount of structure in the data collection process affect saturation? Additionally, with the rapid increase in Internet access and bandwidth over the past decade, qualitative research is experiencing a growth in remote data collection techniques. This raises the question of whether these findings

will hold true for other modalities of collecting focus group data, such as synchronous and asynchronous online approaches. We hope that these data will serve as a foundation for the work of future scholars, and we encourage researchers to examine data saturation in other contexts.

Acknowledgments

Authors' Note

All views expressed in this article, however, are solely those of the authors.

Declaration of Conflicting Interests

Funding

Supplemental Material

The online appendices/data supplements/etc. are available at http://journals.sagepub.com/doi/suppl/10.1177/1525822X16639015.

Notes

1. These percentages have been estimated by scaling data off of the line graph in the authors' book.
2. For the purposes of this article, we consider themes as units of meaning that exist at the nexus between an analyst and the data reviewed. The depth and breadth of themes for a particular study depend on factors including the analytic objectives and the analyst coding style. Codes are explicit labels for, and descriptions of, themes and are formally described in a codebook. Although themes and codes are closely related, from a phenomenological perspective they are different.
3. Saturation analysis was conducted manually based on matrix coding query results. Time and resource constraints prohibited creating more than 10 data sets.

See also Namey et al. (2016) for further analyses of focus group and interview saturation points.

References

Agar, M., and J. MacDonald. 1995. Focus groups and ethnography. *Human Organization* 54:78–86.

Barbour, R. 2007. *Doing Focus Groups*. Thousand Oaks, CA: Sage.

Bluff, R. 1997. Evaluating qualitative research. *British Journal of Midwifery* 5: 232–35.

Byrne, M. 2001. Evaluating the findings of qualitative research. *AORN Journal* 73: 703–6.

Carey, M. 1995. Comment: Concerns in the analysis of focus group data. *Qualitative Health Research* 5:487–95.

Charmaz, K. 2014. *Constructing Grounded Theory*. 2nd ed. Thousand Oaks, CA: Sage.

Cheek, J. 2000. An untold story: Doing funded qualitative research. In *Handbook for Qualitative Research*, eds. N. K. Denzin and Y. S. Lincoln, 401–20. Thousand Oaks, CA: Sage.

Coenen, M., T. Coenen, A. Stamm, G. Stucki, and A. Cieza. 2012. Individual interviews and focus groups in patients with rheumatoid arthritis: A comparison of two qualitative methods. *Quality of Life Research* 21:359–70.

Duggleby, W. 2005. What about focus group interaction data? *Qualitative Health Research* 15:832–40.

Fern, E. F. 1982. The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research* 19:1–13.

Fossey, E., C. Harvey, F. McDermott, and L. Davidson. 2002. Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry* 36:717–32.

Francis, J. J., M. Johnston, C. Robertson, L. Glidewell, V. Entwistle, M. P. Eccles, and J. M. Grimshaw. 2010. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health* 25:1229–45.

Fugard, A., and H. Potts. 2015. Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology*. http://dx.doi.org/10.1080/13645579.2015.1005453 (accessed April 8, 2015).

Galvin, R. 2015. How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *Journal of Building Engineering* 1:1–12.

Glaser, B., and A. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick, NJ: Aldine.

Greenbaum, T. L. 1997. *The Handbook for Focus Group Research*. Thousand Oaks, CA: Sage.

Guest, G. 2015. Sampling and selecting participants in field research. In *Handbook of Methods in Cultural Anthropology*, 2nd ed., eds. H. R. Bernard and C. C. Gravlee, 215–50. Lanham, MD: Rowman & Littlefield.

Guest, G., A. Bunce, and L. Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18:59–82.

Guest, G., and K. MacQueen. 2008. Reevaluating guidelines in qualitative research. In *Handbook for Team-based Qualitative Research*, eds. G. Guest and K. MacQueen, 205–26. Lanham, MD: AltaMira.

Guest, G., K. MacQueen, and E. Namey. 2012. *Applied Thematic Analysis*. Thousand Oaks, CA: Sage.

Hagaman, A., and A. Wutich. 2017. How many interviews are enough to identify metathemes in multi-sited and cross-cultural research? Another perspective on Guest, Bunce, and Johnson's (2006) landmark study. *Field Methods* 29:23–41.

Kaplowitz, M. D., and J. P. Hoehn. 2001. Do focus groups and individual interviews reveal the same information for natural resource valuation? *Ecological Economics* 36:237–47.

Kirchberger, I., M. Coenen, F. Hierl, C. Dieterle, J. Seissler, G. Stucki, and A. Cieza. 2009. Validation of the International Classification of Functioning, Disability and Health (ICF) core set for diabetes mellitus from the patient perspective using focus groups. *Diabetic Medicine* 26:700–707.

Kitzinger, J., and R. S. Barbour. 2001. Introduction: The challenge and promise of focus groups. In *Developing Focus Group Research: Politics, Theory, and Practice*, eds. J. Kitzinger and R. S. Barbour, 1–20. London: Sage.

Knodel, J. 1993. The design and analysis of focus group studies: A practical approach. In *Successful Focus Groups: Advancing the State of the Art*, ed. D. L. Morgan, 35–50. Thousand Oaks, CA: Sage.

Krueger, R., and M. Casey. 2015. *Focus Groups: A Practical Guide for Applied Research*. 5th ed. Thousand Oaks, CA: Sage.

MacQueen, K., E. McLellan-Lemal, K. Bartholow, and B. Milstein. 2008. Team-based codebook development: Structure, process, and agreement. In *Handbook for Team-based Qualitative Research*, eds. G. Guest and K. MacQueen, 119–36. Lanham, MD: AltaMira.

McLellan, E., K. M. MacQueen, and J. Neidig. 2003. Beyond the qualitative interview: Data preparation and transcription. *Field Methods* 15:63–84.

Morgan, D. 1988. *Focus Groups as Qualitative Research*. Thousand Oaks, CA: Sage.

Morgan, D. 1996. Focus groups. *Annual Review of Sociology* 22:129–52.

Morgan, M., B. Fischoff, A. Bostrom, and C. Atman. 2002. *Risk Communication: A Mental Models Approach*. New York: Cambridge University Press.

Morse, J. 1995. The significance of saturation. *Qualitative Health Research* 5: 147–49.

Namey, E., G. Guest, K. McKenna, and M. Chen. 2016. Evaluating bang for the buck: A cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. *American Journal of Evaluation* 37: 425–40.

Powell, R., and H. Single. 1996. Focus groups. *International Journal of Qualitative Health Care* 8:499–504.

QSR. 2012. *NVivo Qualitative Data Analysis Software*. Version 10, 2012. Victoria, Australia: QSR International Pty.

Romney, A., S. Weller, and W. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88:313–38.

Sandelowski, M. 1995. Sample size in qualitative research. *Research in Nursing and Health* 18:179–83.

Ulin, P., E. Tolley, and E. Robinson. 2005. *Qualitative Methods in Public Health: A Field Guide for Applied Research*. San Francisco: Jossey-Bass.

Vaughn, S., J. Shay Schumm, and J. Sinagub. 1996. *Focus Group Interviews in Education and Psychology*. Thousand Oaks, CA: Sage.

Weller, S., and A. Romney. 1988. *Systematic Data Collection*. Newbury Park, CA: Sage.