American Journal of Qualitative Research
2024, Vol. 8 No. 3, pp. 193-207
https://doi.org/10.29333/ajqr/14895

AMERICAN JOURNAL OF QUALITATIVE RESEARCH
ISSN: 2576-2141

n-Sizes, Attributes, and A Priori Sampling: A Qualitative Sampling Model for Large, Heterogeneous Populations

Natalie Perez[1]
Amazon, United States

ABSTRACT
Qualitative sampling in the age of Big Data requires tactful negotiation. Although qualitative research aims to explore the depth as opposed to breadth of experiences, opinions, or beliefs of individuals regarding a unique phenomenon, stakeholders or sponsors might not always be convinced that small sample sizes can yield big results. Intimate population awareness, identification of attributes of importance, selection of a purposeful numbers game, and strategic use of instruments can aid in appropriate sampling approaches for large, heterogeneous populations. This paper reviews the principles of nonprobability sampling, summarizes key qualitative sampling characteristics to consider, and provides a set of examples for negotiating sample sizes in the era of Big Data.

KEYWORDS: Qualitative methods, nonprobability sampling, a priori, employee experience, sample size

The million-dollar question researchers across disciplines must grapple with is estimating the "right" sample size. The concept of "right" in this context relates to adjectives such as acceptable, reasonable, or justifiable rather than without error or accuracy. Although researchers across disciplines and research methodologies must use sampling to yield efficient and valid results (Palinkas et al., 2015; Patton, 2002), qualitative researchers face peculiar challenges with nonprobability sampling strategies. For instance, purposeful sampling strategies are meant to allow researchers to strategically pick participants in order to better understand a phenomenon of interest (Creswell et al., 2007). However, hand-picking participants, especially when working with a large population, often means there are a number of individuals left out of research opportunities. How does a qualitative researcher decide who to include and who should be included or not when examining large, heterogeneous populations? How large can or should a sample be for a technique like maximum variation sampling? How many choices do qualitative researchers need to make to finalize the "right" sample for a study? Sampling in qualitative research is challenging, as numerous trade-offs are required to finalize a recruitment list. Not only does the art of nonprobability sampling become a difficult pursuit, but when faced with stakeholders or sponsors, some researchers may face increased struggles negotiating big data biases in the context of a large population since research can be judged with an omnipresent quasi-quantitative position (Vasileiou et al., 2018). One question I often receive from novice researchers studying large, heterogeneous populations is: What's the "right" sample size for my study? I always remark *it depends*.

Nonprobability sampling strategies offer varying sample sizes, such as 1 participant via an interpretative phenomenological design (Perez, 2023) or over 200 participants via a multi-

[1] Corresponding author; Research Scientist at Amazon. E-mail; natkper@amazon.com

case study design (Brink, 2018), but I have yet to identify literature that illustrates qualitative researchers negotiating large population challenges with a complex, heterogeneous population. What trade-offs does one make in these situations? Of the studies I have reviewed, there is often less focus on a population size, sampling frame, and actual sample size or how demographic attributes can inform and shape a final sampling frame and sample size. Therefore, while I do not seek to leverage this article as an all-encompassing authoritative guide to sampling for qualitative research, my aim is to offer support to guide researchers in navigating how to tackle qualitative inquiry with strategic sampling when faced with large populations and big data opportunities.

This article briefly explains the philosophy of qualitative sampling, a priori sampling theory, research designs, population attributes, and sampling sizes in relation to instruments; it also includes two use cases that provide examples of nonprobability sampling across large populations and ways to negotiate stakeholder perceptions.

A Brief Overview of Nonprobability Sampling

In business, social sciences, ecology, medicine, and most other fields, researchers define two types of sampling strategies: probability sampling and non-probability sampling. Of the two options, probability sampling is largely viewed as the "gold standard" of sampling due to its unbiased characteristics in selecting population quantities (Pescott et al., 2023). In probability sampling, strategies are used to ensure accurate population estimates are represented in a given sample, such as weights addressing measurement error and selection bias (Pescott et al., 2023). In non-nonprobability sampling, samples are often not representative of a larger population, and population estimators are not likely precise (Lincoln & Guba, 1985). However, nonprobability sampling is not inherently bad. Lincoln (1988) argued that non-probability sampling is never representative since representation is not its purpose. Instead, this form of sampling is designed to "exploit competing views" and offer "fresh perspectives as fully as possible" (Lincoln, 1988, p. 273). Qualitative studies focus on achieving a depth of understanding, rather than breadth of understanding like most quantitative studies, about a given phenomenon (Patton, 2022). The focus of depth can vary for qualitative studies, and the amount of participants is arguably one factor influencing how deep or shallow a researcher might explore and interpret perceptions or experiences within the context of a phenomenon (Palinkas et al., 2015; Perez, 2023; Smith et al., 2008; Yin, 2014).

Entering into the era of Big Data, qualitative researchers with access to large populations must navigate inferential complexities using nonprobability sampling techniques (Ary et al., 2018; Creswell, 2013; Palinkas et al., 2015). There are a number of different nonprobability sampling techniques such as criterion-i (Palinkas et al., 2015), criterion-e (Palinkas et al., 2015), typical case (Patton, 2002), homogeneity (Creswell, 2013), snowball (Patton, 2002), extreme or deviant (Smith et al., 2008, intensity (Kramer & Burns, 2008), maximum variation (Bachman et al., 2009), critical case (Patton, 2002), convenience (Creswell et al., 2007), and theoretical sampling (Palinkas et al., 2015), among others. The aforementioned strategies are used purposefully to identify or expand the "variability or dispersion of values for a particular variable or variables, or to narrow down the range of varication and focus on similarities" (Palinkas et al., 2015).

Population, Sampling Frame, and Sample Size

Researchers must first determine their population, regardless of their sampling approach. A population is almost always defined in statistical terms and refers to group of people with parity in defining characteristics (Krieger, 2012). Before considering who to invite to participate in a qualitative study, researchers should pay special attention to the population they wish to study. In particular, researchers must define their study population (Creswell, 2000). When defining a population, researchers should identify key attributes that make up their population of interest, such as geographic location, ethnic/racial group, age, tenure, gender, etc. before selecting a sampling technique (Banerjee & Chaudnury, 2010). For instance, perhaps a qualitative researcher working in the organizational psychology space is tasked with researching Z corporation's employee population; in this context, all 5,000 employees working for Z corporation might be considered a population. However, perhaps this same researcher is asked to study only highly tenured employees who have worked for Z corporation for five or more years. With these additional population attributes, the researcher's population shifts from all 5,000 employees to 478 employees who have five or more years of work experience at Z Corporation. Defining the characteristics of a population is key. According to Krieger (2012), populations can be defined by a number of characteristics or attributes such as linguistic, geographic, socio-political, culture, age, and more (e.g., "elderly population," "Nordic population," etc.).

Once a researcher defines their population and key attributes, it is important to consider a sampling method; that is, after defining an entire group that a researcher might want to draw conclusions about, it is imperative to draw up a sample or more than one sample of a specific group of individuals (Creswell & Garrett, 2008; Palinkas et al., 2015). In particular, researchers might have reason to compare groups or individuals or not compare at all. When considering sampling techniques, researchers might want to look at individuals that make up a particular sub-group to reduce the possibility of variation or identify participants across cases that capture the maximum possibility of variations across important attributes (Palinkas et al., 2015).

After determining whether or not a researcher will conduct a group comparison, a researcher can begin to generate a list of potential individuals to recruit; this is known as a sampling frame. A sampling frame often includes a subset of the population of interest (Mooney & Harber, 2019). A sampling frame usually contains a list of potential participants with some form of contact information that researchers can select potential participants from (Mooney & Harber, 2019); in this way, a sampling frame should be a list of individuals from whom the study's sample will be taken. The number of a population, also known as the *n*-size, will impact the number of individuals in a sampling frame; this is true for both probability and nonprobability sampling, but a sampling frame will have a much lower percentage of potential participants in qualitative research, especially if the *n*-size of the population is large, such as over 500 individuals. A sampling frame can also be impacted by the type of qualitative research design used in a given study. For instance, if an entire population under study contains 106 individuals, a sampling frame for a multi-case study might include all individuals in the population (i.e., 100% of the population), or, if conducting a phenomenological study, perhaps only 30 individuals are necessary to gain the essence of the topic or area of study with respect to the population (i.e., 32% of the entire population). On the other hand, if there is a population of 6,106 individuals, a multi-case study sampling frame might make up 200 individuals (i.e., 12% of the entire population), and a phenomenological study may still only need 30 individuals across a wide range of attributes (i.e., 0.5% of the entire population); hence, the size of the sampling frame is dependent on both the population size and desired qualitative research design.

Since qualitative researchers do not aim to test hypotheses, samples are often significantly smaller compared to quantitative samples (Creswell & Miller, 2000). There is no clear or straightforward answer to selecting a sample size for a qualitative research study (Hennink & Kaiser, 2022). Sample sizes for qualitative research rely on several factors, including epistemology, methodology, and practical issues (Baker & Edwards, 2012; Vasileiou et al., 2018). Scholars such as Yin (2014) indicate that "practical issues" include time, financial constraints, and project deadlines. Yet, practical issues can also include stakeholder beliefs and requests. In fields such as business, medicine, or education, researchers often work closely with stakeholders and sponsors. From experience, it is not uncommon for project sponsors to be unfamiliar with the qualitative research field or the purpose of the depth of inquiry to question small sample *n*-sizes. While many stakeholders and sponsors can gain useful information from researchers, such as education on qualitative research purposes, methodologies, and practices, sometimes education is not enough to convince stakeholders or sponsors that a sample of 12 people is enough to get the information they desire, based on research design and study purpose. In fact, stakeholders or sponsors might hold strong opinions or views regarding Big Data bias – that is, the more individuals, the better the output. These views can be difficult for qualitative researchers, whether novices or not, to overcome.

A Case for a Priori Sampling, Saturation, and Instruments

Several systematic qualitative reviews have explored the use of saturation to estimate sample sizes (e.g., Hennink & Kaiser, 2022). Saturation is an important estimator of qualitative sample sizes, but scholars vary on sample sizes required to reach saturation, mostly based on the degree of homogeneity of the sample, depth of analysis, or research design. The following table offers a range of examples highlighting the interplay between study population, sample size, saturation, and research design.

Table 1
Articles Illustrating Sampling, Design, and Saturation

Author	Research Design	Population	Sample Size	Saturation
Charmaz (2006)	Grounded theory	Homogenous	25	Higher-order groupings
Yin (2009)	Single-case study	Homogenous	25-50	N/A
Stake (2005)	Multiple-case study	Heterogenous	4-10 *(cases; individuals unknown)*	N/A
Smith et al. (2008)	Interpretative phenomenological study	Homogenous	3-10	Codes and categories; higher order groupings
Coenen et al. (2012)	Basic interpretative qualitative study	Homogenous	39	Higher-order groupings
Hagaman and Wutich (2017)	Multiple-case study	Heterogenous	132 *(20-40 per site)*	Codes and categories
Young and Casey (2018)	Meta-analysis	Heterogenous	27	Codes and categories

Qualitative research focuses on the meaning, not just the occurrence or frequency, of real-life situations, events, or phenomena (Yin, 2011). While qualitative researchers are less concerned with frequency, larger sample sizes with repeatable patterns can aid with increased confidence in a dataset as well as benefit researchers in capturing a range of perceptions, experiences, or opinions. Greater participant diversity can lead to richer insights and allow researchers to identify patterns, contradictions, nuances, and unusual cases that lead to a greater understanding of the richness and context of people. If larger sample sizes can benefit a study, and if heterogeneous samples require more participants than homogenous populations, how many participants are needed for a qualitative study with a heterogeneous population? *It depends.* The good news is there are some helpful steps to work through this complex process.

An A Priori Model

There are several approaches for determining sample size, including a rule-of-thumb approach based on previous studies (Sim et al., 2018), conceptual models based on study characteristics (e.g., Malterud et al., 2016; Morse, 2000), numerical estimates (e.g., Ando et al., 2014; Francis et al., 2010; Guest et al., 2006; Hennink et al., 2017), or statistical formulas (e.g., van Rijnsoever, 2015; Tran et al., 2016). Most a priori studies reference resources, project deadlines, or time constraints as some reasons for determining a qualitative sample size upfront rather than throughout the process (Sim et al., 2018). However, Sim et al. (2018) argued that many a priori models can be questionable or inappropriate given the nature of qualitative methodologies. Instead, they suggest that a preferable approach is to generate a provisional sample size with an anticipated upper limit without generating a precise prediction (Sim et al., 2018). Adopting these recommendations, this article reinforces the benefits of adopting a "loose" a priori sampling model. By "loose," we argue that researchers should determine the anticipated limits, both minimum and maximum amount of potential participants for a given study. This practice can be a value-add to researchers working in fields such as business, where aggressive study deadlines might make continuous sampling flexibility difficult. To determine the sample limits of a large, heterogenous population, researchers should consider three key elements: study design, demographic attributes, and saturation thresholds.

Study Research Design

Qualitative research designs that might focus more on the breath of information compared to depth usually contain a greater sample size (e.g., basic interpretative qualitative case-studies), whereas qualitative designs that rely on multiple coding and analytical rounds that illicit meaning across multiple viewpoints might require fewer participants (e.g., narrative, phenomenology). There are also research designs that can joist between breadth and depth (e.g., longitudinal, grounded theory), but these studies might contain more participants than other studies. The depth of analysis is an important feature to consider when assessing the number of participants needed for a given study, and this is consistent with heterogeneous populations as it is with homogeneous populations.

Like proponents of conceptual models of a priori sampling, I argue that the purpose and research design should inform its sample size scope. Consider an example from the educational field, where a team of researchers is tasked with identifying student perceptions of a 4th-grade nationwide curriculum. The research team decides to use a basic interpretative design, and they want to include voices from a variety of fourth-graders. Hypothetically speaking, the population of 4th-graders in the United States includes a large number of individuals. These individuals likely have a range of diverse upbringings, experiences, and perceptions about their educational experiences, particularly the nationwide 4th-grade curriculum they are using. When considering potential attributes, the researchers might include geographical location, gender, socioeconomic family status, etc. The research team plans to use maximum variation sampling to ensure a wide range of diverse participant viewpoints are included. Identifying key attributes is a critical mechanism for the research team to determine not only the adequacy of data or evidence required for their study (Erickson, 1986) but also to support selecting the types of individuals who should participate and be represented within the study.

The research team begins to examine their population first by reviewing statistics from the Department of Education to determine the average 4th-grader attributes. Through a quick scan, the research team determined that there are more 4th-graders enrolled in urban schools than in rural schools. The male/female ratio is 50%, and most 4th-graders live within a family means of $45,000-$60,000. From an initial glance, examining the average attributes might seem like a basic way to pick "typical" cases of 4th-graders across the United States. However, diving deeper, one member of the research team found that there were far more nuances in defining "urban schools," and students had very different lifestyles across states with an average family income of $45,000-$60,000. When trying to dive deeper and define attributes and the range of attributes to consider for a given population, researchers might find this practice can become increasingly complex, which is why it is important to determine the type of attributes and extent of attributes researchers deem critical to understanding a phenomenon. Using contextual and statistical insights, the research team decides on several key attributes (i.e., some demographic attributes, academic records, and class attendance records) to aid the team in dividing the heterogeneous population of 4th-grade students across the U.S.

Attributes and Saturation

When working with large, heterogenous populations, attributes are important to stratifying a population into groups. In probability sampling, one common sampling technique is stratified sampling. This type of sampling divides a heterogeneous population into homogenous groups or sub-groups, also known as strata, generally based on one or more attributes (Elfil & Negida, 2017). The benefit of this type of sampling approach is that it allows researchers to gain details on effect size while also obtaining samples from minority or under-represented people groups (Elfil & Negida, 2017). The concept of "effect size" is akin to the strength of a statistical claim, whether or not a researcher can be confident in the research results. Borrowing a similar strategy, qualitative nonprobability sampling can benefit from identifying and exploring individual attributes and compiling sub-groups or homogenous groups that can be paired against saturation thresholds to determine whether or not a finding might be consistent for a particular group but not the whole or whether or not a finding might be consistent across groups and therefore inform findings for the whole sample.

Sim et al. (2018) issue a word of caution to qualitative researchers when deciding on sample sizes by arguing that methodological knowledge is key to determining how saturation and other relevant parameters might inform the requirements of a study, yet contextualization is essential to informing sample sizes with numerical guidelines. For heterogeneous studies, there is no numerical sample size or guideline to follow, but rather, the sample size depends on research purpose, design, and attribute splicing. Nevertheless, this study proposes a type of homogenizing of a heterogeneous population to aid in systematic analysis and saturation as an aspect of sampling and analysis. Researchers can homogenize a heterogeneous population through a range of attributes or based on context from a range of factors. For instance, recall the researcher tasked with studying tenured employees from Z corporation. This researcher was told that there were some anecdotes that managers experienced their work very differently than non-managers. This context is important, and the researcher must consider whether or not the factor of manager status is worth differentiating. In the end, when faced with over 400 employees, the researcher splits the heterogenous population by the employee attribute of manager status (i.e., non-managers and managers). Using the employee attribute of manager status, the researcher divided the the tenured population into two segments: managers (n=98) and non-managers (n= 380). The researcher then generated a sampling frame for managers and a different sampling frame for non-managers.

Splicing a population into two or more groups based on one or more attributes can help researchers better identify whether or not there are any meaningful differences within and across groups a part of a larger population, as well as aid in determining saturation and data confidence assessment across groups. However, researchers should also be careful to avoid overlapping individuals across subgroups. Importantly, though, with more attributes and groups, it means more analysis time and effort; researchers must consider the trade-off in the number of sub-groups they wish to create from a heterogeneous population.

Saturation has been scrutinized in various ways across studies and scholars in fields such as medicine and social sciences (e.g., Hennink & Kaiser, 2022; Saunders et al., 2018). The concept of saturation has several meanings, such as coding saturation and meaning saturation (Hennink & Kaiser, 2022), which suggests the ways researchers operationalize and achieve saturation can vary. When considering saturation in terms of sampling, it refers to assessing the adequacy of a sample size and also the point of data collection when no new issues or insights are identified, rendering the data repetitive (Hennink & Kaiser, 2022). The means to achieve saturation range from empirical to statistical strategies (Hennink & Kaiser, 2022), but less is known about how parameters can influence saturation. By incorporating attributes and homogenizing heterogeneous populations within a qualitative study, the use of saturation can take on a different purpose, aside from meaning or coding repeition. From this vantage point, researchers can also use saturation to gauge whether or not there is enough data to adeuaqtely compare within groups and across groups. Saturation is also contingent on the degree of variability and the phenomenon of interest (Hennick et al., 2017). In other words, if people have largely similar experiences, saturation will be reached more quickly than if there is a wide variety of experiences or perceptions.

Homogenizing Heterogenous Populations

The following section illustrates two hypothetical examples that consider ways to homogenize heterogenous populations. While the examples posed in this section are nearer to the field of organizational psychology, researchers engaging in qualitative sampling across any field might yield from the principles and lessons learned from the use cases. The underlying premise of the use cases is to aid researchers in identifying design and sampling decisions that can influence a study's outcomes.

Use-Case 1. Attributes and Saturation

In the first use-case researchers are tasked with conducting a mixed-methods sequential explanatory study to explore perceptions of organizational culture among employees living around the world. During the first phase of the study, data is drawn from a census survey that measures cultural perceptions of employee's work environment. In the second phase, researchers use criterion-i sampling (i.e., select all cases that meet some predetermine criteria) and stratified purposeful sampling (i.e., forming mostly homogenous sub-groups from a heterogenous population). With a population of 2,000 individuals, the team decides to examine attributes that might inform cultural perceptions based on two attributes that presented differences based on the survey analysis: geographical region and organizational division. These two attributes are used to cluster the 2,000 individuals in the population into six groups (i.e., 3 regions, 2 divisions). The researchers use focus groups to collect the qualitative data. They rely on previous numerical models (i.e., Stake, 2005) as a guide to determine a maximum sampling frame (i.e., 4-8 focus groups per attribute). The researchers build a series of focus groups per region first, then by division with a total of 10 invitees per group. The minimum number of focus groups planned is 24, and the maximum planned is 48. The minimum desired number of participants is 72 (i.e., minimum of n=3 individuals per focus group), and the desired maximum number of participants is 480 (i.e., maximum of n=10 individuals per focus group). The sampling frame for the study includes 480 individuals; this number reflects the researchers' impressions of the size and diversity of the total population based on the selected attributes.

After the study is conducted, the final sample size includes 25 focus groups with 79 participants. Diving into the data, the research team uses bracketing and analyzes the data one region at a time and one division at a time. After all divisions and regions are analyzed from within-groups, the team moves to across-group analysis and examines the entire dataset.

The researchers are compelled with the findings. While they gained a general sense of cultural perceptions and nuances across regions, when they attempted to delve deeper into country-specific aspects within each respective region, they found that data saturation was lacking. *Why*? The researchers did not collect enough data across participants within each country housed underneath a specific region to identify whether or not country-level differences were based on unique, typical, or by-chance aspects related to the phenomenon. In this way, while the attribute of geographic region was valuable and reached coding and meaning saturation (i.e., frequency counts, coding, categories, and higher order themes) from a country-level, there was insufficient data to reach saturation. Aside from regional findings, the researchers could not gain adequate findings across countries embedded in the regions, since they did not use that attribute to guide their sampling frame. In addition, when looking across organizations, the researchers did not identify meaningful cultural differences across divisions.

Use-Case 1. Lessons Learned. There are several lessons learned from this use-case example. The first is that sampling can influence saturation. The level of sampling at a regional level attribute meant that saturation could likely only be achieved at a regional level. Since not enough individuals participated within any specific country, a deeper-level of saturation could not be attributed from within a country-level perspective. Secondly, sometimes the attributes selected to homogenize a heterogenous population do not reveal meaningful differences; this is not inherently bad, it simply means that, in this context, organization might not inform cultural differences for individuals working across a global organization. Perhaps a deeper-level of inquiry into specific departments within an organization might have revealed differences or not, but the study did not explore any department specifically, as it did not explore any country specifically. Thirdly, this study contained a large number of participants, but it could have included fewer or more individuals. Using a priori strategies, the researchers determined a multi-case study research design was appropriate given the study's purpose and objectives. The researchers determined criterion-i sampling was acceptable given the nature of the two attributes identified in the survey research (i.e., respondents were selected based on 1/ participation in the survey, 2/ geographic location, and 3/ organizational division). The research~~er~~ team invited over 400 participants but deemed 72 individuals across two organizations and regions to be the minimally acceptable number of participants to achieve the aims of this study. To justify the sampling decision, the researchers determined: 1/ the breadth of participant voices was acceptable based on conceptual and numerical models used to loosely generate a sampling frame (i.e., looking for broader insights, less deeper insights), 2/ two attributes were deemed important by stakeholders and researchers, 3/ coding and meaning saturation based on the attributes used to splice the population into groups (i.e., regional level and organizational division level) were deemed acceptable, given the study's focus and desired outcomes. While 79 individuals participated, as few as 72-participants would likely have enabled researchers to meet the a priori participation sampling estimates and achieve saturation for the study's aim, which was to explore organizational culture from the perspectives of employees around the world.

Use-Case 2. Scaffolding Qualitative Analysis through Homogenous Bracketing

In this use-case researchers conduct an interpretative qualitative study aimed at understanding employees' perceptions of meaningful work. The employees are clustered in a region that includes eight countries. Four business leaders want to receive insights into their employee's perception of meaningful work. Consequently, researchers determined that a deeper-level of analysis was needed per leader. Since four business leaders were a part of the study, the researchers used business leader as an attribute. In addition, there were contextual anecdotes that non-manager experiences were different compared to manager experiences.

Hence, two attributes were used to homogenize the region's heterogenous population: leader directs and manager status. Since the researchers do not have access to potential connections between attributes and meaningful work, they want to ensure a wide-range of employees and characteristics are captured in the study. The researchers opt to use maximum variation sampling. To aid in maximum variation sampling, the researchers identify several secondary attributes including country, gender, employee tenure, and job family. Overall, the primary attributes are business leader and manager status, and the secondary attributes are country, gender, employee tenure, and job family. Importantly, the researchers will ensure saturation is reached at the leader level and manager level but not the secondary attributes. *Why*? Because the aim is to deliver insights to business leaders pertaining to their employee's perceptions of meaningful work across managers and non-managers, and saturation will be evaluated from based on those attributes. The secondary attribute's purpose is to aid researchers identify all possible ranges of viewpoints across each leader's employee population. In this way, the researchers recruit individuals to participate in their study first based on leader and manager status, and then based on a wide range of characteristics (e.g., tenure, job family, etc.), in an attempt to achieve the widest range of employee perceptions as possible.

After identifying the population and selecting key primary and secondary attributes, the researchers decide on the minimum and maximum sampling frame requirements. First, the researchers identify the total homogenous population per leader; each of the business leaders in the study had a total population of roughly 100 employees (10 managers/90 non-managers). Given the range in sample sizes across manager and non-manager status, the researchers opted to use focus groups to collect data from the non-managers, while in-depth interviews were used to collect data from the managers. Using loose numerical estimates (i.e., Andres, 2022; Yin, 2014) the researchers aim to use 4-8 focus groups for non-managers per business leader (i.e., minimum of *n*=3 participants and maximum of *n*=8 participants desired) and at least 6 in-depth interviews for managers per business leader (e.g., scholars suggest 6-12 in-depth interviews; Creswell, 2009). Across all leaders, the researchers aim to conduct a minimum of 16 focus groups and maximum of 32 focus groups as well as a minimum of 24 in-depth interviews (i.e., *n*=6 managers per leader) and a maximum of 40 in-depth interviews (i.e., *n*=10 managers per leader). The researchers determined that the minimum required number of participants for non-managers is 48 and the maximum number of non-managers is 256, and the minimum number of managers is 24 and the maximum number of manager is 40. In addition, the minimum number of participants per leader is 18 employees (i.e., 4 focus groups with *n*=3 non-managers; *n*=6 manager interviews), and the maximum number per leader is 74 employees (i.e., 8 focus groups with *n*=8 non-managers; *n*=10 manager interviews). The sampling frame for the study included 74% of each business leader's respective employee population (i.e., maximum: *n*=74 per leader; *n*=296 across all four leaders).

Overall, 122 individuals participated in the study (i.e., *n*=98 non-managers/*n*=24 managers). The researchers used bracketing and analyzed one business leader's entire dataset at a time; each business leader had roughly 23-24 non-managers and 6 managers participate in the study. Analyzing each leader's dataset at a time made the analytical process less overwhelming, as the research team was able to conduct multiple rounds of coding to engage in deeper-analysis and interpretation per leader and manager status. Using frequency counts, several rounds of different coding strategies (e.g., in vivo, values, etc), categorization, and higher order themes, the researchers determined that data saturation was met for each leader and manager status. After analyzing within groups, the researchers analyzed findings across groups. Within leader groups, managers and non-managers showed little variability in their responses related to meaningful work. However, across leaders, there was greater variability in employee perceptions. In other words, the researchers found that when examining data within cases, there was higher rates of saturation and lower rates of variance; however, when analyzing data across cases, saturation was lower. While secondary attributes were not used to gauge saturation, they offered a way range of employee characteristics that allowed researchers to hone in on key concepts addressed across a wide range of employees. The business leader attribute was revealed to be a key attribute, with high rates of variance and less saturation from a cross-group analysis compared to a within group analysis. Consequently, the researchers not only discovered insights related to unique aspects within each leader's employee population regarding meaningful work, but also found that employee's perceptions of meaningful work ranged considerably when comparing employee experiences across business leaders but less across manager status.

Use-Case 2. Lessons Learned. This study revealed several lessons learned including: 1/ dividing the heterogenous population by two attributes was key to achieving the degree of depth desired by researchers and stakeholders, 2/ secondary sampling attributes were helpful when using maximum variation sampling to identify and select a wide range of employee characteristics, 3/ managers and non-managers showed little difference in perceptions of meaningful work within groups, but greater differences in perceptions across leader groups, 4/ big *n*-sizes are less cognitively overwhelming when grouped into smaller strata for analysis, since breadth of findings was important for business leaders for this study.

Differences and Similarities Across Use-Case 1 and Use-Case 2

The use-cases differ in their purposes, employee populations, research questions, and research designs. The first use-case had a sampling frame nearly double that of the second, due to differences in the depth and breadth of the research. Different attributes were also used across the two use-cases, and the sampling approaches varied as well. Despite these differences, the use-cases shared several similarities. Both aimed to determine key attributes used to homogenize heterogeneous populations, and both used saturation as a key indicator of the quality of findings. Additionally, the studies employed similar data-collection strategies with "loose" numerical guidelines to aid in determining the appropriate range for the minimum and maximum required sampling frame to likely reach saturation.

Overall, the use-cases promote the use of a priori strategies to estimate population sampling in order to help researchers determine the most adequate number of respondents. This approach, while not always typical in the qualitative field, may particularly benefit researchers working in settings with tight deadlines or limited flexibility to expand data-collection before analysis. Researchers should consider their time, resources, and project constraints to determine if this approach is beneficial for their work. The following section offers a linear process to support qualitative researchers working with heterogeneous populations.

Not a Net-New Nonprobability Heterogenous Sampling Model

With few defined sampling models or practices for qualitative researchers to use when navigating large, heterogenous populations, this section offers researchers a systematic model to use for both sampling and analytical estimation. Although this work does not promote a net-new sampling model, it does pull together a mixture of sampling techniques into a qualitative model that defines and uses attributes, sampling frames, and saturation to group and examine heterogenous populations:

1. Determine population characteristics: Examine, where possible, details about the population such as "averages" or "extremes" through external literature, internal data, or stakeholder insights;
2. Calculate the population size and diversity: Ensure the population size is accurately calculated and the diversity of possible attributes in the population is taken into account (e.g., location, age, gender, etc.);

3. Determine breadth and depth of analysis: Determine the extent of coding analysis, including the types of coding techniques and the number of analytical rounds of coding, as well as the within cases compared to across case needs;
4. Identify meaningful attributes: Select attributes, whether by study focus, theory, context, stakeholder, and/or researcher judgment, to homogenize the heterogeneous population. Researchers can generate primary and secondary attributes, but primary attributes should be used to inform subgroup creation and saturation estimation;
5. Define population subgroups by attributes: Individuals should be placed into one subgroup based on one or more attributes. There should not be individuals overlapping across subgroups. Subgroups in nonprobability heterogeneous sampling might include a different allocation of sizes (e.g., one subgroup might contain n=60 participants, while a different subgroup might contain n=12 participants). Ensure the allocation size of the subgroup contains enough individuals to achieve saturation via loose sampling estimates depending on instrument;
6. Generate sampling frame: A sampling frame should include an a priori sample of the minimum and maximum number of potential participants, given the subgroup attributes and study objectives. Previous conceptual or numerical models regarding saturation (e.g., 9-17 for in-depth interviews or 4-8 focus groups, according to Hennink & Kaiser, 2022) can be valuable starting points for determining a plausible numerical range of participants per subgroup and the intended instrument(s) used for data collection;
7. Finalize sample size: Recruit participants. Use the sampling frame to assess whether or not the minimum desired sample size has been reached across each subgroup;
8. Ensure saturation is reached: The point of redundancy is important to achieving coding saturation, as data analysis continues until no new information or nothing new is heard in the cases (Patton, 2022). Where possible, the flexibility of sampling should unfold, especially if seeking to identify discomforting or negative cases (Miles & Huberman, 1994), but project or timeline constraints may restrict the extent to which a researcher can engage in flexible sampling. Meaning and coding saturation should be achieved across each subgroup and associated attribute(s). If saturation is not achieved, more participants might be needed.

Recalling Sim et al. (2018), this model is suggested to be used loosely, as it involves segments of conceptual or numerical models that are deconceptualized from any one qualitative study. That said, this model reinforces that attributes are one way to use study parameters to estimate saturation from nonprobability sampling. Overall, population sizes, diversity, and meaningful attributes, as well as saturation, require tailored sample sizes for each study's sampling frame and final sample size.

Conclusion

Estimating sample sizes for qualitative research is inevitably a numbers game. Large populations, heterogenous in nature, require a wider range of participants to ensure dimensions of variability of constructs and associated attributes are accounted for within or across groups for a given population. Sample sizes should be large enough to account for new and diverse understanding (Sandelowski, 2000). In nonprobability sampling, researchers often but not always prioritize depth over breadth and seek participants who can offer rich information about a given phenomenon (Creswell et al., 2008). However, qualitative studies can also obtain rich information from larger datasets. The diversity of a population is a valid concern for qualitative researchers, and researchers must negotiate sample sizes and trade-offs to determine at what point no further disconfirming evidence can be found (Baum, 2002).

Borrowing principles from probability sampling, particularly stratified sampling, can aid qualitative researchers in narrowing large, heterogeneous populations into purposeful groupings. These groupings can not only aid researchers in better understanding the diversity of a population, but they can also aid researchers' practicality by chunking datasets into groups or cases that invite deeper case-by-case analysis before analyzing cross-case datasets.

In their classic work, Guba and Lincoln (1981) argued that nonprobability sampling is not meant to be random but purposeful; the aim of this approach is to "exploit competing views" and obtain "fresh" viewpoints as wide as possible (p. 273). From this position, diversity in viewpoint is a benefit, and collecting a range of viewpoints from individuals across a population might have a relatively stable numerical saturation point, as identified from previous studies, and qualitative researchers are encouraged to expand opportunities to ensure a spectrum of viewpoints are collected, documented, and analyzed when working with large, heterogeneous populations.

References

Ando, H., Cousins, R., & Young, C. (2014). Achieving saturation in thematic analysis: Development and refinement of a codebook. *Comprehensive Psychology*, *3*, Article 4. https://doi.org/10.2466/03.CP.3.4

Ary, D., Jacobs, C. L., Irvine, C. K. S., Walker, D. (2018). *Introduction to Research in Education*. Cengage Learning.

Bachman, M. O., O'Brien, M., Husbands, C., Shreeve, A., Jones, N., Watson, J., Reading, R., Thoburn, J., & Mugford, M. (2009). Integrating children's services in England: National evaluation of children's trusts. *Child: Care, Health and Development*, *35*(2), 257–265. https://doi.org/10.1111/j.1365-2214.2008.00928.x

Baker, S. E., Edwards, R., & Doidge, M. (2012). *How many qualitative interviews is enough?: expert voices and early career reflections on sampling and cases in qualitative research*. National Centre for Research Methods, Southampton. https://research.brighton.ac.uk/files/301922/how_many_interviews.pdf

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. *Industrial Psychiatry Journal*, *19*(1), 60–65. https://doi.org/10.4103/0972-6748.77642

Baum, M. A. (2002). The constituent foundations of the rally-round-the-flag phenomenon. *International Studies Quarterly*, *46*(2), 263–298. https://doi.org/10.1111/1468-2478.00232

Brink, R. (2018). A multiple case design for the investigation of information management processes for work-integrated learning. *International Journal of Work-Integrated Learning*, *19*(3), 223–235. https://files.eric.ed.gov/fulltext/EJ1196745.pdf

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Coenen M, Stamm TA, Stucki G, Cieza A. (2012). Individual interviews and focus groups in patients with rheumatoid arthritis: a comparison of two qualitative methods. *Qual Life Res*. 21(2):359-70. https://doi.org/10.1007/s11136-011-9943-2.

Creswell, J. W. (2013). *Steps in conducting a scholarly mixed methods study*. DBER Speaker Series, 48. https://digitalcommons.unl.edu/dberspeakers/48

Creswell, J. W., & Garrett, A. L. (2008). The "movement" of mixed methods research and the role of educators. *South African Journal of Education*, *28*(3), 321–333. https://doi.org/10.15700/saje.v28n3a176

Creswell, J. W., Hanson, W. E., Clark Plano, V. L., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist*, *35*(2), 236–264. https://doi.org/10.1177/0011000006287390

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Elfil, M., & Negida, A. (2017). Sampling methods in clinical research; an educational review. *Emergency*, *5*(1), Article e52. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5325924/

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). McMillan.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health*, *25*(10), 1229–1245. https://doi.org/10.1080/08870440903194015

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. Jossey-Bass.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hagaman, A. K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on Guest, Bunce, and Johnson's (2006) landmark study. *Field Methods, 29*(1), 23–41. https://doi.org/10.1177/1525822X16640447

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, *292*, Article 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, *27*(4), 591–608. https://doi.org/10.1177/1049732316665344

Kramer, T. L., & Burns, B. J. (2008). Implementing cognitive behavioral therapy in the real world: A case study of two mental health centers. *Implementation Science*, 3(1), 1–10. https://doi.org/10.1186/1748-5908-3-14

Krieger, N. (2012). Who and what is a "population"? Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. *The Milbank Quarterly*, *90*(4), 634–681. https://doi.org/10.1111/j.1468-0009.2012.00678.x

Lincoln, Y. S. (1998). The ethics of teaching in qualitative research. *Qualitative Inquiry*, 4(3), 315–327. https://doi.org/10.1177/107780049800400301

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research,* 26, 1753–1760. https://doi.org/10.1177/1049732315617444

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Mooney, S. J., & Garber, M. D. (2019). Sampling and sampling frames in big data epidemiology. *Current Epidemiology Reports,* *6*(1),14–22. https://doi.org/10.1007/s40471-019-0179-y

Morse, J. M. (2000). Determining sample size. *Qualitative Health Research*, *10*(1), 3–5. https://doi.org/10.1177/104973200129118183

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*, 533–544. https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. SAGE Publications.

Perez, N. (2022). *“Who am I?”: An interpretative phenomenological analysis of tutors’ lived experiences providing online writing tutoring* [Doctoral dissertation, University of Hawai'i at Manoa]. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/54e2b3bc-5e77-4e3a-90f7-326b9a139ce7/content

Perez, N. (2023). The rebirth of online tutoring from desert to oasis: An interpretive phenomenological analysis illustrating a writing tutor’s journey transitioning online through the stages of grief. *The Peer Review Journal, 7*(3). https://thepeerreview-iwca.org/issues/issue-7-3/the-rebirth-of-online-tutoring/

Pescott, O. L., Boyd, R. J., Powney, G. D., & Stewart, G. B. (2023). *Towards a unified approach to formal risk of bias assessments for causal and descriptive inference*. Airxv. https://doi.org/10.48550/arXiv.2308.11458

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340. https://doi.org/10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Sim, J., Saunders, B., Waterfield., J. & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodolog*y, *21*(5), 619–634. https://doi.org/10.1080/13645579.2018.1454643

Smith, J. A., & Osborn, M. (2008). Interpretive phenomenological analysis. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (pp. 53–80). SAGE Publications.

Stake, R. E. (2005). Qualitative case studies. In: N. K. Denzin and Y. S. Lincoln (eds.). The SAGE handbook of qualitative research (3rd ed.). Thousand Oaks, CA: Sage, pp. 443–466.

Tran, L. N., Back, A. L., & Creutzfeldt, C. J. (2016). Palliative care consultations in the neuro-ICU: A qualitative study. *Neurocritical Care*, *25*, 266–272. https://doi.org/10.1007/s12028-016-0283-5

van Rijnsoever, F. J. (2017). (I can’t get no) saturation: A simulation and guidelines for sample sizes in qualitative research. *PloS One*, *12*(7), Article e0181689. https://doi.org/10.1371/journal.pone.0181689

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*, 1–18. https://doi.org/10.1186/s12874-018-0594-7

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

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed). SAGE Publications.

Young, D. & Casey, E. (2018). An examination of the sufficiency of small qualitative samples. *Social Work Research. 43*(1). https://doi.org/10.1093/swr/svy026

Notes on Contributor

Natalie Perez, Ph.D. works as a Research Scientist at Amazon. She has a unique interdisciplinary background that integrates educational technology, retention support services, instructional design, and organizational psychology with an emphasis on mixed methods and design-based research within higher education and corporate environments. Her secondary line of research focuses on qualitative methodologies and generative artificial intelligence (GenAI), including its use and application, particularly within the context of qualitative research and data analysis.

ORCID

Natalie Perez, https://orcid.org/0000-0003-3981-0468