A choir or cacophony? Sample sizes and quality of conveying participants’ voices in phenomenological research

Theodore T Bartholomew¹, Eileen E Joy², Ellice Kang² and Jill Brown³

Abstract
Misunderstandings about qualitative methods, whether phenomenological or otherwise, are prevalent in social science research. Such misunderstandings leave researchers, reviewers, and editors less equipped to conduct or evaluate this method. Evaluation of phenomenology is especially complicated given the different variants that exist and the need for flexibility within these studies. Methodologists have created guides for conducting specific variants of phenomenology; however, these do not provide clear guidance as to what is an adequate sample in phenomenology. The purpose of this systematic review was to help improve implementation of phenomenological methods by exploring sample issues as they relate to study quality. We implemented an explanatory sequential mixed methods design to test relationships between samples and studies’ quality then deepen our understanding of these findings with a focused content analysis. First, we reviewed and coded 200 manuscripts following the PRISMA method. Larger samples were associated with lower quality and studies aligned with a specific phenomenological method tended to be of higher quality. Second, we identified two cases from the studies reviewed and subjected them to deductive qualitative content analysis to identify features that demonstrate quality. Findings are discussed with respect to implications for phenomenological methods in social and health sciences.

Keywords
Phenomenology, sample sizes, sample types

Phenomenological research methods, regardless of their type, are important to empirical approaches in psychology, social, and health sciences that center thick descriptions of lived experiences, the voices of individuals who share an experienced phenomenon, and the meaning of a specific phenomenon. These methods emerge from phenomenological philosophy (Husserl, 1980 [1911]), with later approaches stemming from Heidegger’s (1962) work. Adapted for use in research methods, variants of phenomenological research, including descriptive (Colaizzi, 1978; Giorgi, 2009), interpretive (Smith et al., 2009), transcendental (Moustakas, 1994), and hermeneutic (van Manen, 1990, 2014) phenomenological approaches have emerged in social sciences, education, and other fields. Shared across these variants, however, is an assumption that meaning can be phenomenologically reduced from participants’ expressed consciousness. Participants in these studies, though each has unique lived experiences, share having lived the phenomenon under study. Sample sizes in these approaches, however, are not standardized and confusion persists with respect to understanding appropriate sample sizes in phenomenology and qualitative research. In conjunction with a mainstream approach in fields like psychological science that prioritizes post-positivism, quantitative methods, and large sample sizes for generalizations, a lack of standards for samples in phenomenology poses problems for researchers emphasizing science that relies on small samples in fields that prioritize larger sizes.

¹Department of Psychology and Department of Africana Studies, Scripps College, Claremont, CA, USA
²Department of Educational Studies, Purdue University, West Lafayette, IN, USA
³Department of Psychology, Creighton University, Omaha, NE, USA

Corresponding author:
Theodore T Bartholomew, Department of Psychology and Department of Africana Studies, Scripps College, 1030 Columbia Ave., Claremont, CA 91711, USA. Email: tbarthol@scrippscollege.edu
The need for identifying standard samples in phenomenology is not akin to the typical flaws ascribed to qualitative methods by quantitatively biased researchers (i.e. that phenomenological samples are too small for generalizability). These methods do not target generalization or development of universal truth. However, a “research method has to be a procedure that can be implemented by many” (Giorgi, 2009: 112), and this need for clarity includes a sense of how many participants is sufficient in phenomenological research. Theorists have generally discussed the concept of sample sizes in qualitative methods broadly (Blaikie, 2018; Conlon et al., 2020; Malterud et al., 2016; Sim et al., 2018); however, this body of work also directs researchers’ attention to the unique sample needs of analytical methods (cf. Conlon et al., 2020; Malterud et al., 2016). Editors, reviewers, and researchers across fields will benefit from a shared understanding of what is common with respect to samples in phenomenological studies. With this missing information, the capacity to evaluate this empirical approach may improve for reviewers who would otherwise impose post-positivist or other qualitative ideas onto their evaluation of phenomenological work. That is, by empirically examining sample sizes within phenomenological research, those who may typically apply standards of post-positivist paradigms to qualitative evaluation may instead embrace the particulars of phenomenological samples. Thus, the purpose of our study was to identify the typical samples in phenomenological research studies through a systematic mixed methods review, associate these sample sizes and types with a measure of study quality, and explore quality in unique cases as a means to better understand what is commonly done regarding sample sizes in phenomenological studies.

**Variants of phenomenological research**

The philosophical traditions of Edmund Husserl and Martin Heidegger informed the development of a diverse range of phenomenological research methods, which assume that reality is subjective and based on individual experiences and meaning (Giorgi et al., 2017). The following two overarching schools of phenomenology exist: descriptive phenomenology based on Husserl’s teachings and a more interpretative approach drawn from Heidegger’s work (Finlay, 2012; Reiners, 2012). Husserl argued that by intentionally setting aside preexisting ideas, conscious experiences can be described without bias (Reiners, 2012). Heidegger (1962) altered the focus of phenomenological study from perceiving to being and dismissed the need to set aside preexisting ideas for interpreting a phenomenon. A range of phenomenological methods applied in accordance with Husserl’s or Heidegger’s philosophical assumptions have emerged whereas others have suggested a movement beyond an essence-oriented approach (e.g. Vagle’s post-intentional phenomenology; Vagle, 2018; Vagle and Hofseß, 2015). Most popular are the following:

(a) descriptive which aims to systematically describe the structure of a phenomenon as it is reflected in the lived experiences of participants using language that reflects the transformation of participants’ natural expressions into psychological expressions (Giorgi, 2009);

(b) transcendental, defined by Moustakas (1994), which describes the structure of a phenomenon as purely as possible, emphasizing the need for researchers to set aside preexisting beliefs;

(c) hermeneutic which provides researchers with a tool to interpret unique lived human experiences through introspection by “borrow[ing]” (van Manen, 1990: 62) experiences of others to understand a phenomenon or lived identity;

(d) interpretative which, though similar to hermeneutic, prioritizes a more idiographic focus and the “details” and “particulars” associated with experiences, events, and emotions (Smith et al., 2009: 29).

**Common threads between phenomenological methods**

Although these types of phenomenological research are unique, they share critical features such as adopting a phenomenological attitude that facilitates imaginative variation (i.e. the process of identifying the essential features of a phenomenon; Finlay, 2008) and an emphasis on the expressions of a few individuals. Adopting a phenomenological attitude is reflected in researchers’ seeing a phenomenon newly and approaching participants’ lived experiences with curiosity, openness, and genuineness (Finlay, 2014). Doing so primes researchers to lean in closely to participants’ experiences, dwell with the minutiae of their experiences, pull narratives together to understand meaning, and express the meaning of participants’ consciousness after hearing their voices (Finlay, 2014).

Taking on a phenomenological attitude and deriving meaning from expressed consciousness warrants considering from whom data are collected. The variants of phenomenology most commonly espouse the collection of individual data, often interviews but also other forms of expressed consciousness, from participants. Focusing on the individual allows researchers to center individual expressions and draw shared understandings across participants. This is congruent with the purpose of phenomenology—to embrace the details of an individual’s expressed experiences, describe them and understand their meaning rather than explore how those experiences are conveyed or what dialectical exchange influences those expressions. Others have discussed the use of focus groups in phenomenological data collection (Bradbury-Jones et al., 2009; Palmer et al., 2010); however, both individual and group approaches are unified by the belief that nuanced understandings of individuals emerge through a person’s expressions. These emphases on unique expressions and imaginative variation inherently influence...
sampling in phenomenology. That is, too many voices could foster discord, or a cacophony, when depicting results whereas a focused sample may encourage a more harmonious expression of participants’ interpreted lived experiences in which individual voices are honored and can contribute to a coherent whole.

**Participant samples in phenomenological research**

Rarely do formalized methods of phenomenological research identify sufficient samples, with the interpretative approach as the only variant suggesting possible sample sizes (Smith et al., 2009). Smith and colleagues provide some opaque direction: (a) Master’s level research projects could be limited to three participants, (b) doctoral dissertations are more difficult to quantify with respect to sample sizes, (c) between three and six participants “can be reasonable for a student project” (Smith et al., 2009: 51), and (d) sample sizes are best if they range from 4 to 10 interviews. These authors also state, despite these guiding comments, that “no right answer [exists] to the question of the sample size” (Smith et al., 2009: 52). Guidelines predicated on career stage influencing sample size and an uncertain standard for what appears to be more professional research leaves this guide less clear than it may initially appear.

Additional texts in descriptive (Giorgi, 2009), hermeneutic (van Manen, 1990), and transcendental (Moustakas, 1994) phenomenology do not specify an adequate number of participants. Wertz (2005) extended this lack of clear direction further, suggesting that sample sizes are pragmatic in that they are tailored to the research questions being asked, “The question of ‘how many participants?’ can only be answered properly by considering the nature of the research problem and the potential yield of findings” (p. 171). Giorgi (2009) offered more abstract direction for sample sizes, indicating that a single person is not enough but that each participant’s voice must be heard in the presentation of results: “Data will have come from multiple persons rather than a single individual. However, the unity of the consciousness of each person has to be respected” (p. 102). Respecting this unity of consciousness may take shape in using pseudonyms or alternative names for individuals interviewed rather than simply attributing quotations to “participants.” Extracting from these methodologists, it appears one person is sufficient if it adequately answers a research question, but with several participants, the researcher must also ensure the voice of each participant is respected.

However, overview texts are frequently relied upon in qualitative studies as they offer some additional direction. Creswell and Poth (2017), in a description of phenomenology, pointed to Polkinghorne’s (1989) suggestion that phenomenological studies ought to contain participant samples between 5 and 25. Perhaps this suggestion from Polkinghorne (1989) is the only clear-cut suggestion or range of sample sizes in phenomenological methods. This, however, is in contrast with some of the variant specific recommendations, or lack thereof.

Despite this room for empirical misunderstanding, variants of phenomenological methods do point to a need for samples to be small. For instance, interpretive phenomenological analyses are typically “conducted on relatively small sample sizes which are sufficient for the potential of IPA to be realized” (Smith, 2011: 10). Authors of phenomenological methods have suggested the issue of saturation (i.e. dealing with redundancy; Wertz, 2005), the need to fully voice participants’ expressed consciousness (Giorgi, 2009), and the inability to adequately present participants’ quotations (Smith, 2011) as reasons small samples are necessary. However, constructs like saturation do not themselves require a small sample and definitions of saturation, whether conceptualized theoretically or the exhaustion of data, differ across qualitative methods leading to variable application of saturation by researchers (see Malterud et al., 2016). These assertions impart the sense that too much data in phenomenological methods risks suffocating or blurring the voices of the participants, thereby distancing the results from their expressions of consciousness. The tenuous dance between the harmony of the choir and the cacophony of rogue solo performers is the dance that phenomenological researchers do when embarking on a study, often times without guidance. Furthermore, phenomenological methods exist in a social science context where generalization is ascribed great value, where biases toward certain methods may lead to the imposition of post-positivist sample size standards on phenomenological work (van Manen, 2014). This assertion is no less true of sample standards of other qualitative methods (e.g. grounded theory) being imposed on phenomenological methods. These biases (i.e. preferences for large samples) could inhibit the publication and influence of phenomenological research. Unfortunately, and even though these approaches put forth a theoretical argument for small samples, those unfamiliar with phenomenology and qualitative methods are still unlikely to have a clear sense of what is acceptable with respect to phenomenology sample sizes.

**The current study**

With the intention of offering greater understanding of phenomenology samples, the aim of our research was to systematically explore the samples (sizes and types) used in phenomenological studies and test connections between samples, types of phenomenology, and study quality. Others have undertaken similar systematic reviews and analyses of methods in qualitative studies (e.g. Woods et al., 2016), supporting the usefulness of empirically examining steps in qualitative research. As can be seen, methodologists offer minimal clarity with respect to the question “how many participants is enough.” Our intent is not to dictate how many are enough but, rather, assert that samples can be more clearly understood in phenomenology to the benefit of
researchers, consumers, editors, and reviewers. Suggesting a priori standards of sample sizes in qualitative research is likely counter-productive (Sim et al., 2018) as such decisions need to be made based on theory and the ability of the available data to answer a given research question. Although sample decisions should be driven by study needs, it remains valuable to communicate to stakeholders in empirical fields what samples might look like.

We also contend that the quality of phenomenological manuscripts is necessary to consider when clarifying commonly used sample sizes and variants in phenomenological research. This extends from methodologists’ assertions, if vague at times, that phenomenological analyses require attention to individuals’ expressed consciousness. Thus, it follows that well conducted phenomenological studies would likely have a small sample and use individual data sources. To address our purpose, we used an explanatory sequential mixed methods approach (Creswell and Plano Clark, 2017) with priority given to the quantitative phase. This began with systematically identifying studies that indicated use of phenomenological methods available in a research database and subsequently employing a quantitative measure of quality. This was followed with a qualitative content analysis intended to offer some additional understanding to our exploration of the relationship between samples and quality (quality from Lockwood et al., 2015 described in the ‘Method’ section). Our mixed methods approach is born of a pragmatic paradigm, which supports the simultaneous use of qualitative and quantitative methods of inquiry to generate practical evidence to support best practice. Although quantitatively studying a qualitative method may seem counterintuitive and an imposition of postpositivist ideals, a quantitative approach can aid in sharpening understanding of what is commonly done within a method. Moreover, concepts across disciplines point to the power of mixing methods. For instance, an etic (scientist-oriented) approach shifts the focus from local observations, categories, explanations, and interpretations to those of the social scientist. The etic approach realizes that members of a culture often are too involved in what they are doing to interpret their cultures impartially. Alternatively, the emic approach investigates how local people think (Kottak, 2006) and how they perceive and categorize the world. We begin with a more “etic” approach by using a view from above with our systematic review and quantitative analysis followed by attention to the “emic” or within-study characteristics.

Furthermore, systematically reviewing and coding phenomenological studies can offer useful insight into common research practices and help to align with Giorgi’s (2009) assertion that even flexible methods like phenomenology should be procedural so it is implementable by many. Our quantitative phase, however, is descriptive and exploratory in nature and is not intended to discount the assertion that sample sizes in qualitative methods ought not be determined a priori (Sim et al., 2018). By integrating these quantitative findings with qualitative content analysis, we are able to apply what is learned in a quantitative phase with an explanatory qualitative phase. We set out to test the following hypothesis with our data:

- Sample size will be negatively associated with rated study quality such that an increase in sample size will predict a decrease in study quality.

In addition, because our process is exploratory in nature, we focused on the following research questions:

- What are the average sample sizes of phenomenological studies across specific phenomenological methods and is there a significant difference?
- Does phenomenological studies’ quality differ by type of phenomenology used?
- Does study quality vary by the type of sample (i.e. individual participants or focus groups)?
- What features can be coded in cases drawn from the review to help understand quality within phenomenological manuscripts in light of quantitative findings about samples?

**Method**

**Studies sampled**

The phenomenological studies were selected based on a systematic review process (described below). All articles were identified using a PsycINFO search, which resulted in the initial identification of 1754 articles. We opted to review 200 of these articles as this allowed us adequate statistical power (see ‘Analysis strategy’ section). Articles came from various peer reviewed journals indexed in PsycINFO. To be included in the sample, the authors must have identified that a phenomenological method was used for the study. This meant studies that may have misnamed their work were included; however, this was done so that we could see what is commonly conducted under the name of “phenomenological research” which also would afford a better understanding of quality. We included all variants of phenomenology research methods.

**Reviewers**

The research team consisted of four members (two faculty, two graduate students). All authors have used descriptive and/or transcendental phenomenology in their research. None of the authors’ published phenomenological studies were included in this review. The two graduate students coded all 200 articles as a pair.

**Measures**

**Study quality.** To assess the quality of each study reviewed, we utilized the Joanna Briggs Institute Checklist for Qualitative Research (JBI-CQR; Lockwood et al., 2015). Typically
used to determine the quality of qualitative studies because of its coherence (Hannes et al., 2010), we utilized this scale to provide a numeric quality rating of each article. Other researchers have calculated scores from the checklist (Ehlers et al., 2018; Nilsson et al., 2015; Wride and Bannigan, 2018); however, other studies have not used the checklist in statistical testing. We made alterations to the scale that allowed us to specifically address unique quality components in phenomenological research. The scale consists of 10 items, and we added four additional phenomenology-specific items. Examples from the scale in its original format include “Is there congruity between the research methodology and the methods used to collect data?” “Are participants, and their voices, adequately addressed?” and “Is there congruity between the research methodology and the research questions or objectives?” Additional phenomenology-specific questions included “Have the researchers identified the specific type of phenomenology?” “Is there a unifying synthesis or written essence for the themes?” and “Is there a section of the manuscript devoted to researcher reflexivity or epoché?” These latter two questions are points of debate in some variants of phenomenological methods; however, they are common to several approaches to phenomenological research and were therefore included as quality criteria. Although it is a shared concept in many qualitative methods, we also added the question “Did the researchers describe the sufficiency of their sample size?” which allowed us to reflect on whether or not the authors attended to concepts like saturation or general allusions to sample size sufficiency. Respondents using the scale select “Yes,” “No,” or “Unsure” to identify whether or not the quality indicator is present in the article. To derive a quality score, we scored “Yes” = 1, “Unsure” = 0, and “No” = −1; thus, higher scores reflected higher quality. We found the JBI-CQR had adequate internal consistency using Cronbach’s alpha (Coder 1: \( \alpha = .70 \); Coder 2: \( \alpha = .75 \)).

**Procedure**

We followed the evidence based Preferred Reporting Items for Systematic Reviews and Meta-Analyses standards (PRISMA; Liberati et al., 2009; see Figure 1) to select articles for review. We restricted the search to articles within the PsycINFO database. Search terms included “phenomenological method” or “phenomenological research” or “phenomenological research design” and results were filtered to only include peer reviewed empirical articles. Using these criteria, 1754 articles were identified.

We used two methods of sampling. First, 75 articles by order of relevance (i.e. how relevant the search engine believed the articles identified were to the search terms) were collected to ensure that the current state of phenomenological literature would be reviewed, as over 90% of the articles identified through this search were published within the last decade. Order of relevance was also determined by term frequency, whether search terms were included in the title or abstract, and exact field matching (EBSCOhost, 2018). Articles were included if they stated using phenomenological methods and were accessible, written in English, peer reviewed, and not a mixed methods study. Seventeen articles not meeting these criteria were excluded, leaving 58 articles for review. Second, we randomly selected the remaining articles to ensure the review encompassed a range of the phenomenological literature. Articles were randomly selected until a final sample of 200 was reached.

Each article was independently coded by two researchers to identify variant of phenomenology, sample size, type of sample, use of pseudonyms, and study quality. To improve the reliability of coder ratings, the coders collaboratively rated the first 20 articles until a shared agreement was developed. For the remainder, coding was independent.

After coding and completing quantitative analysis, we looked to find two exemplars in the set of reviewed studies. Identification of particular cases is an advocated for approach in mixed methods research (see Creswell and Plano Clark, 2017). We identified contradictory cases so that we could explore, with greater specificity, the quality of phenomenological studies in light of sample sizes. Specifically, we isolated studies that had a sample size that was above the mean but also had a quality rating that was at least one standard deviation above the mean. We also then identified studies with below average sample sizes but had quality ratings that were at least one standard deviation below the mean. We picked one article from each of these groups as cases that were then subjected to qualitative content analysis.

**Analysis strategy**

**Quantitative.** We first calculated an intra-class correlation coefficient (ICC) as an index of interrater reliability. Specifically, we calculated a Two-Way Mixed ICC because the
same two raters rated each manuscript. We then conducted (Tables 1 and 2) ANOVA tests to determine if sample sizes and study quality were significantly different across variants of phenomenology (descriptive, hermeneutic, interpretive, transcendental, other) and types of sample (individual, focus group, both, other). Subsequently, we analyzed the relationships between sample size and study quality by type of phenomenology using simple correlations (see Table 3). Finally, we used multiple regression to predict study quality. Type of phenomenology, sample size, use of pseudonyms (as a means of expressing individual voices), and type of sample were entered into the model as predictors (see Table 4). Power analysis, conducted using G*Power (Faul et al., 2007), which affords researchers a chance to enter desired power and number of predictors used when predicting outcomes of various statistical models (e.g. $F$ value in the case of multiple regression). This power analysis indicated a minimum sample of 166 to achieve a medium effect size with .95 power for the regression model. We opted for a slightly larger sample size.

**Qualitative.** Thereafter, we identified two cases (large sample, high quality; small sample, low quality) for qualitative content analysis to conduct a manifest analysis of the two

### Table 1. Means, standard deviations, and mean difference tests.

| Type of phenomenology (n) | $M$  | $SD$  | Range   | Model statistics |
|---------------------------|------|-------|---------|------------------|
| Sample size               |      |       |         |                  |
| None reported (n = 62)     | 26.73| 46.39 | 2–308   |                  |
| Descriptive (n = 43)       | 12.14| 3.78  | 1–31    | $F$ (df)         |
| Hermeneutic (n = 38)      | 12.61| 6.14  | 3–28    |                  |
| Other (n = 13)             | 16.00| 16.20 | 3–65    |                  |
| Transcendental (n = 3)     | 8.00 | 2.65  | 5–10    |                  |
| Overall (N = 200)         | 16.87| 27.38 | 1–308   |                  |

Study quality

| Type of phenomenology (n) | $M$  | $SD$  | Range   | Model statistics |
|---------------------------|------|-------|---------|------------------|
| None reported (n = 62)     | 0.74 | 3.73  | −10–9   |                  |
| Descriptive (n = 43)       | 6.28 | 2.78  | −3–11   |                  |
| Hermeneutic (n = 38)      | 6.00 | 1.95  | 1–9     |                  |
| Other (n = 13)             | 4.15 | 3.13  | −1–10   |                  |
| Transcendental (n = 3)     | 10.33| 2.08  | 8–12    |                  |
| Overall (N = 200)         | 4.23 | 3.90  | −10–12  |                  |

LSD: Least Significant Difference; MSE: Mean Square Error; SD: standard deviation.

Reflects the minimum difference required between means in order to identify significance. Other includes identifying the phenomenological approach as feminist phenomenology, a phenomenological life-story approach, Spiegelberg’s phenomenology, and so on.

*p = .03; **p < .001.

### Table 2. Type of sample and study quality means, standard deviations, and ANOVA results.

| Type of sample (n) | $M$  | $SD$  | Model statistics |
|-------------------|------|-------|------------------|
| Study quality     |      |       |                  |
| Individual (n = 179) | 4.61 | 3.70  |                  |
| Focus group and individual (n = 9) | 0.89 | 3.33  | $F$ (df)         |
| Focus group (n = 7) | 2.43 | 4.12  | 6.53** (3, 196)  |
| Other (n = 5)     | −0.80| 5.63  |                  |
| Overall (n = 200) | 4.23 | 3.90  |                  |

LSD: Least Significant Difference; MSE: Mean Square Error; SD: standard deviation.

Reflects the minimum difference required between means in order to identify significance. Other includes written responses, journaling, photographs, and so on.

**p < .001.

### Table 3. Correlation between sample size and study quality by phenomenology type.

| Type of Phenomenology (n) | $R$  | $p$   |
|---------------------------|------|-------|
| None reported (n = 62)     | −.46 | <.001 |
| Descriptive (n = 43)       | −.12 | .44   |
| Hermeneutic (n = 38)      | −.21 | .20   |
| Other (n = 13)             | .04  | .82   |
| Transcendental (n = 3)     | −.10 | .75   |
| Overall (N = 200)         | −.40 | <.001 |

Thereafter, we identified two cases (large sample, high quality; small sample, low quality) for qualitative content analysis to conduct a manifest analysis of the two
identified studies. We opted for these unique cases because it would allow us to explore instances when quality was contrary to expectations set by the quantitative results. For this supporting phase in the study, we used a deductive approach to qualitative content analysis (Elo and Kyngas, 2007). Because our quantitative phase involved an existing structure of study quality, the preparation phase began with the complete review of the two manuscripts selected. Organization followed through compiling evidence supporting the quality coding for the manuscripts into the categorization matrix. That is, we organized meaning units (or those aspects of manuscripts salient to the quality/sample of a study) in accordance with the quality indicators used in the quantitative phase. The first author coded the two manuscripts, and the fourth author reviewed and confirmed the content analysis.

**Results**

Prior to exploring our questions, we determined that the interrater reliability of the raters was strong, ICC = .948. Results indicated most frequently studies did not report the type of phenomenology used (n = 62). This was followed by studies using a descriptive (n = 43), interpretive (n = 41), hermeneutic (n = 38), other (n = 13), or transcendental (n = 3) approach. The 13 studies categorized as other did identify a type of phenomenology; however, these were not variants with accompanying texts to guide the process. These included feminist, life-story approach, Spiegelberg’s phenomenology, and other identifiers. We opted to group these together as they were identified by the researchers but did not fit with the other variants. Most studies in the sample utilized individual participant interviews (n = 179), whereas fewer used focus groups and individual interviews (n = 9), focus groups alone (n = 7), or other sources of data such as journals or photographs (n = 5). In addition, 63 of the studies used pseudonyms for participants. Looking at a specific item in our adapted study quality scale, we found that just 18 (9.0%) studies sampled discussed the sufficiency of the sample size (n = 47, 23.5% of studies were unclear in sample size discussion; n = 135, 63.5% of studies did not discuss sample size sufficiency at all). We also found that few studies included a section on researcher reflexivity or epoché (n = 10, 5%) based on a phenomenology-specific quality item we included in the quality measure. Mean values of variables are reported in Tables 1 and 2.

**Sample size differences by phenomenology approach**

For the studies sampled, the average sample size was just over 16 (SD = 27.38; range = 1–308). Omnibus ANOVA results indicated a significant difference in sample sizes across the variants of phenomenology identified (see Table 1). Using the least significant difference value (i.e. the minimum difference between two means that indicates a significant difference) for pairwise comparison, we identified that significant differences were only observed between studies that did not report the type of phenomenology used and those that did report a specific type (descriptive, interpretive, hermeneutic, or transcendental). That is, studies without an identified type of phenomenology, on average, had larger sample sizes (see Table 1) than studies that were identified as using descriptive, interpretive, hermeneutic, other, or transcendental approaches. The effect sizes of these differences were moderate, with Cohen’s $d$ values (see Cohen, 1988)

### Table 4. Multiple regression model with study quality as the dependent variable.

| Variable                  | $\beta$ | SE  | 95% CI       | t    | p     |
|---------------------------|---------|-----|--------------|------|-------|
| Sample size               | -.03    | .01 | -.05, -.02   | -3.86| <.001 |
| Type of phenomenology     |         |     |              |      |       |
| Descriptive (n = 43)      | 4.89    | .59 | 3.73, 6.05   | 8.32 | <.001 |
| Interpretive (n = 41)     | 3.87    | .60 | 2.69, 5.05   | 6.48 | <.001 |
| Hermeneutic (n = 38)     | 4.69    | .60 | 3.51, 5.88   | 7.81 | <.001 |
| Other (n = 13)            | 2.95    | .90 | 1.17, 4.72   | 3.28 | .001  |
| Transcendental (n = 3)    | 8.69    | 1.70| 5.34, 12.05  | 5.11 | <.001 |
| Type of sample            |         |     |              |      |       |
| Individual (n = 179)      | .89     | 1.12| -1.30, 3.09  | .43  | .425  |
| Focus group and individual (n = 9) | -1.52  | 1.44| -4.37, 1.34  | -1.05| .296  |
| Other (n = 5)             | -.75    | 1.74| -4.19, 2.69  | -.43 | .669  |
| Pseudonyms                | .55     | .46 | -.36, 1.46   | 1.203| .23   |
| Constant                  | .81     | 1.15| -1.47, 3.09  | .70  |       |

$\beta$ = standardized regression coefficient. For Type of Phenomenology, “None Identified” (n = 62) is the reference group. For Type of Sample “Focus Groups” is used as the reference group. Pseudonyms are a binary variable with no pseudonyms or names used for participants as the reference group.
ranging from .40 to .70. There were no additional significant differences in this model.

**Study quality differences by phenomenology approach**

Average quality rating for the studies was 4.23 (SD = 3.90), with a range of −10–12. An ANOVA test revealed an overall significant mean difference in study quality by the types of phenomenology used (see Table 2). Pairwise comparisons indicated that studies without an identified type of phenomenology in the manuscript were, on average, of less quality than all others. Descriptive, interpretive, hermeneutic, and transcendental studies were all also of significantly higher quality than were those manuscripts categorized as other (see Table 1 for descriptive statistics). The effect sizes (i.e. Cohen’s $d$ ranging from 1.49 to 3.10) for these comparisons were exceptionally large as the mean for the non-identified phenomenology group was multiple standard deviations lower than the means of other categories. Transcendental studies, though few, were significantly higher quality than all other categories. No significant differences in quality were observed between descriptive, interpretive, and hermeneutic studies.

**Study quality differences by type of sample**

Because what constitutes a sample in phenomenology is also less clear in theoretical writings about these methods, we tested study quality by type of sample. Type of sample included individual participants, focus group and individual participants, focus groups, and other samples (e.g. journal writings). ANOVA results, reported in Table 2 with descriptive statistics, demonstrated an omnibus significant difference in study quality across the types of sample. Pairwise comparisons demonstrate that those samples with only individual participants were of significantly higher quality than all other sample types. Effect sizes for these comparison’s ranged from moderate (i.e. Individual samples vs Focus Group Samples: $d = .71$) to very strong (i.e. Individual vs Both Focus Group and Individual: $d = 1.21$; Individual vs Other Sample: $d = 1.70$). Focus group samples were also of significantly higher quality than were Other samples (Cohen’s $d = 1.05$) and samples of combined focus groups and individuals (Cohen’s $d = .50$).

**Relationship between samples and study quality**

Before testing the regression model to answer our hypothesis, we tested the bivariate relationships between sample size and study quality by phenomenological approach. We found only significant relationships for transcendental phenomenological studies ($r = .99$, $p = .03$) and for those studies without an identified phenomenological approach ($r = −.46$, $p < .001$). The transcendental correlation coefficient should be considered with extreme caution as this is calculated from just three studies. However, among those studies where authors do not identify the type of phenomenology, larger sample sizes were related to lower study quality. Sample size in the remaining phenomenological variants was not related to study quality (see Table 3).

Subsequently, we tested a multiple regression model with study quality as the outcome variable. Sample size, type of phenomenology, type of sample, and the use of pseudonyms were entered as predictors. The model (see Table 4) significantly predicted study quality, $F(10, 189) = 18.09$, $MSE = 2.87$, $R^2 = .49$, $p < .001$. Results demonstrate that sample size significantly predicts study quality ($β = −.03$, $p < .001$). Thus, increases in sample size are associated with decreases in study quality when type of phenomenology, type of sample, and the use of pseudonyms are held constant. Type of phenomenology was also significant in the model. In this model, the reference group was those studies that did not report a type of phenomenology ($n = 62$). The coefficients indicate that studies identified as descriptive, interpretive, hermeneutic, other, or transcendental were, on average, significantly higher quality than studies without an identified phenomenological approach when the other variables were held constant. However, type of sample and use of pseudonyms were not significant predictors of study quality.

**Two unique cases: large sample size, high quality; small sample size, low quality**

We identified two studies to conduct our manifest qualitative content analysis. Our focus was on what was done within these manuscripts that contributed to the quality interpretation. Extreme, unique cases were chosen (high quality, large sample and low quality, small sample) so we could better contextualize our quantitative results (i.e. studies with larger samples were more likely to be of lesser quality). The large sample, high quality study had a sample of 22 participants and was agreed upon by the raters as having a 9 ($M = 4.23$, $SD = 3.90$, possible range = −14–14) quality score. The small sample, low quality study had a sample of three participants and a quality score of −3. The large sample, high quality paper collected data through individual interviews and used descriptive phenomenology. Similarly, the small sample, low quality used individual interviews ($N = 3$) and emails from participants was identified as a descriptive phenomenological study. Neither study used pseudonyms for participants. Sample size does not always dictate that the voices of participants are present. In the small sample, low quality study, participant voices were lacking and only a minimal use of significant statements and meaning units was evident. The analysis more broadly captured themes without the depth common to phenomenology. This analysis also revealed more information about the complex relationship between sample size and quality. In the large sample, high quality study, the researchers explicitly discussed their
choice in sample size, citing that the sample was large enough to extrapolate themes, structure, and essence. This was not mentioned in the small sample, low quality study, leading us to wonder if the mere conscientiousness of sample size might predict quality. Despite sample size being an important aspect of phenomenological methods, it operates in collaboration with other aspects of the study to predict quality. In Table 5, we present the content analysis results to demonstrate facets of each article that contributed to the interpretation of quality.

Discussion

Our purpose with this systematic mixed methods review was to develop clarity regarding common sample sizes and types in phenomenological studies identified in a predominantly psychological, though interdisciplinary, search engine. In the hopes of seeing the forest for the trees, we linked this to an index of study quality to put forth an empirically based understanding of how samples are related to overall quality of phenomenological studies. This approach involves some reduction of quality within a nuanced method to a statistical value. We offer an indication of what might be typical phenomenological samples identified in PsycINFO. Furthermore, we add to the ongoing discussion about types of sample in phenomenological methods (Bradbury-Jones et al., 2009; Palmer et al., 2010) and sample sizes broadly in qualitative inquiry (Blaikie, 2018; Conlon et al., 2020; Malterud et al., 2016; Sim et al., 2018). We use these findings to speculate about sample sizes and types in phenomenology, the need for adherence to methods associated with phenomenology variants, the implications of phenomenology for health research, psychology, and related fields, and recommendations for implementing phenomenological methods.

Adequate sample sizes in phenomenological research

As noted, the variants of phenomenological methods pose different, though vague, concepts of what constitutes a sufficient sample. A lack of direction can be confusing for researchers attempting to implement best-practices in phenomenological research and for reviewers, editors, and readers tasked with weighing the rigorous conduct of a study. With the data from the current study, we have demonstrated that sample size is, overall, related to study quality such that larger sample sizes coincide with lower quality research articles. For instance, we coded a study that was identified as phenomenological research in which researchers described collecting data from 308 participants, and we ultimately rated this study to be of very poor quality (rating score = −10; more than four standard deviations below the mean). Thus, we speculate that larger samples suppress the voices of participants and descriptions of phenomena, inhibiting the expression of details and deep experiences in analyses, and are overall poorly suited to phenomenological research. This appears to especially be the case when researchers claim to use phenomenology but fail to align themselves with a specific variant.

Attention to correlations between sample size and study quality by phenomenology variant helps to clarify the influence of our findings. Sample sizes in the transcendental studies (n = 3) were correlated to study quality, with larger samples (range = 5–10) being associated with higher quality; however, this correlation coefficient is calculated from a much smaller sample and confined range of sample sizes. In addition, we did not find correlations between sample sizes in the three most common phenomenology variants (descriptive, interpretive, and hermeneutic) and study quality. We take this to indicate that studies conducted in accordance with the standards of these approaches can be of sufficient quality at, near, or below the averages for each variant, demonstrating the practical flexibility in sample size decisions within the method. This is supported by our identification of a high quality descriptive phenomenological study that contained an above-average sample size for the qualitative content analysis. Furthermore, we identified no significant differences between average sample sizes in descriptive, hermeneutic, interpretive, or transcendental studies. We also see that small samples do not always result in high quality research, as demonstrated by the small sample (n = 3), low quality manuscript in our qualitative phase. From this, smaller samples may be preferred, but such studies must also adhere to other method guidelines to result in rigorous phenomenological research. This evidence demonstrates that sample sizes may differ across studies to fit unique needs, as is suggested in the theoretical literature (Malterud et al., 2016), and ought to be iterative in nature (Sim et al., 2018). Our data also lead us to assert that researchers using phenomenology should collect a sample small enough to adequately describe expressions of consciousness and center participants’ voices, a sentiment that aligns with predominant phenomenological methods.

Types of samples: are focus groups congruent with phenomenology?

Understanding sample sizes also requires attention to what constitutes a sample in phenomenological research. Most guidance to phenomenological methods supports using individual interviews as data. However, debates indicate some researchers are open to using focus groups (Bradbury-Jones et al., 2009; Palmer et al., 2010). The majority of the manuscripts reviewed here (n = 179) relied on individual interviews. That fewer studies used focus groups (n = 7), focus groups and individual interviews (n = 9), or other data (n = 5) demonstrates a trend to use individual interviews. We found quality differences between the types of samples, such that individual interview samples were of higher quality than all other sample types. Although these analyses were
Table 5. Organization matrix for qualitative content analysis.

| Quality indicator | Interpreted meaning units and ratings | Rating | Supporting meaning units | Rating | Supporting meaning units |
|-------------------|--------------------------------------|--------|--------------------------|--------|--------------------------|
| Large sample (N = 22), high quality (9) | Small sample (N = 3), low quality (−3) |        |                          |        |                          |
| 1. Is there congruity between the stated philosophical perspective and the research methodology? | 1 | • Moving away from cause and effect | 1 | • Epistemological stance connected to individual consciousness |
| |   | • Researchers valued lived experiences and descriptions in context |   | • Belief that data come from people’s expressions |
| 2. Is there congruity between the research methodology and the research question or objectives? | 1 | • Purpose is about meaning | 1 | • Purpose to reveal structure and meaning of phenomenon |
| |   | • Stated the method is suited to the purpose |   |   |
| 3. Is there congruity between the research methodology and the methods used to collect data? | 1 | • Use individual interviews | 1 | • Yes; uses individual interviews |
| |   | • Interviews focus on individual experiences |   | • Clear effort to center the descriptions and lived experiences of participants |
| |   | • Purposive sampling of those who shared experience |   |   |
| 4. Is there congruity between the research methodology and the representation and analysis of data? | 1 | • Cite phenom. methods with respect to analysis | −1 | • Results not thoroughly presented |
| |   | • Describe rigor and verification |   | • Minimal participant quotations included |
| 5. Is there congruity between the research methodology and the interpretation of results? | 0 | • Unclear in manuscript | −1 | • No attention to meaning in results |
| 6. Is there a statement locating the researcher culturally or theoretically? | 0 | • Unclear in manuscript | −1 | • None provided in manuscript |
| 7. Is the influence of the researcher on the research and vice-versa addressed? | 1 | • Identified researcher in table | 1 | • Researcher influence expressed |
| |   | • Described researcher as part of the analytical process |   | • Though expressed, not identified or clarified |
| 8. Are participants and their voices adequately represented? | 1 | • Several quotations throughout results | −1 | • No quotations included in manuscript |
| |   | • Quotations come from multiple participants |   |   |
| 9. Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data? | 1 | • Interpretations remain focused on participants’ experiences | −1 | • Results presented in general descriptions |
| |   | • Discussion does not go beyond scope of data |   | • Voices of participants absent or reduced to minimal quotations |
| |   |   |   | • Overinterpretation of data (i.e., discussion seems biased) |
| 10. Have the researchers identified the specific type of phenomenology? | 1 | • Yes; descriptive phenomenology identified | 1 | • Yes; descriptive phenomenology identified |
| 11. Is there a section of the manuscript devoted to researcher reflexivity or epoche? | −1 | • Appears in table, but not explicitly reflexivity or epoche | −1 | • Mentions bracketing but provides no details |
| 12. Is there a unifying synthesis or written essence for the themes? | 0 | • Summative section at beginning of discussion | −1 | • None provided |
|   |   | • No clear mention of synthesis or essence |   |   |
| 13. Did the researchers describe the sufficiency of their sample size? | 1 | • Report exhaustiveness of data | −1 | • Sample sufficiency not discussed |
| 14. If type of phenomenology is identified, is the analysis congruent with that specific type? | 1 | • Yes; follows aligned analytical approach | 0 | • Uses concepts from multiple types of phenomenology |
| |   |   |   | • Not specific to descriptive phenomenology |

Note: Items 1–9 come from the quality rating scale used for the study. For scoring, as stated in the “Method” section, “Yes” = 1, “Unsure” = 0, and “No” = −1.

*Item added by the authors.
conducted with small samples in the focus group category, they allude to an important point for researchers to consider. That is, is the use of focus groups congruent with a phenomenological emphasis on being close to participants’ expressions of consciousness?

Our data do not directly answer this question, but they add to the conversation. This sample included few studies that used focus groups. Possibly, this is a function of most approaches implying data should come from individual perspectives. It is true that authors sometimes do not explicitly state how participants should be interviewed. For example, Todres and Galvin (2008) wrote, “in traditional phenomenological research, we interview a number of people” (p. 576). This “number of people” is not specified as individually or in groups, but we can extract from the phenomenological philosophy that what is intended is understanding drawn from individuals (cf. Heidegger, 1962; Husserl, 1980 [1911]). If we consider how particular phenomena are conveyed as an individually driven process, then researchers, logically, will prioritize individuals’ expressions.

Others may, however, value a shared expression of lived experiences that may come from group data collection. Vagle, in retelling communication with a colleague, exclaimed “of course, one cannot separate subject and object!” (Freeman and Vagle, 2013: 728), reflecting Husserl’s assertion that subject (i.e. the individual) and object (i.e. the experience) are inseparable. Groups may be a different subject—one that creates different conditions under which individuals express experiences. When a group is a subject, researchers may be grappling with a shared presentation of experience, pieced together and co-constructed by voices that speak in a group. Possibly, the use of a focus group creates a new dynamic of subject and object but could be done in ways that conflate ideas about individuals as subjects. This would, however, misalign with Merleau-Ponty’s (1962 [1945]) conceptualization of intersubjectivity, or an understanding of self with others, within phenomenology. Considering this debate overall, alongside our finding that studies using focus groups were below average quality, it would appear these studies may have flaws. Researchers may also be curious about how a specific group expresses experiences of an object, in which case, focus groups may be congruent to a phenomenological purpose. However, research is needed, beyond theoretical discussions, to determine if focus groups fit phenomenological research. Defining appropriate samples and data sources is further complicated by conceptualizing how an individual should express consciousness. That is, how can researchers best use photography, journaling, or other individual expression in phenomenology? Our data show that such studies tend to also be of lower quality, but this also requires additional research.

Coherence between variant and method

Whether or not focus groups are appropriate for phenomenological research is very much in accordance with the need to develop coherence in the conduct of phenomenological research. Coherence, in research methods, can be thought of as “thoughtfulness concerning the empirical claims made by researchers and whether they fit with the approach and methods taken” (Holloway and Todres, 2003: 345). Researchers have an obligation to identify how they conduct empirical efforts so results are intelligible and communicable. This is true of phenomenological research without abandoning flexibility or creativity in this family of methods. Instead, researchers who believe phenomenology will help answer research questions can conduct their work in thoughtful, coherent ways that connect to phenomenology.

We cannot enter the mind of the researchers who conducted these studies, but our data support the assertion that researchers who simply identify their work as phenomenology without following guidelines of specific approaches ignore or exclude key method considerations. Plausibly, these researchers inaccurately carried out a qualitative study under the guise of phenomenology. This is also true of researchers who report using one phenomenological approach but do not clearly follow it, as was identified in the qualitative content analysis (see Table 5). These findings reinforce Giorgi’s (2009) assertion that methods require specific procedures many can follow. This is further substantiated in our finding that those studies implementing the specific phenomenological variants (descriptive, hermeneutic, interpretive, or transcendental) were of higher quality than those studies who reported a specific but non-guided phenomenological approach (i.e. the “other” category).

Is bigger better or is less more: implications for research

Better understanding the implementation of phenomenological methods has clear implications for improved quality of such work within health care and psychological research. Wertz, for example, summarized eight reasons for which phenomenology fits well with empirical aims in counseling psychology. Among these reasons was that phenomenology is formalized and methodologists have developed “justification and norms concerning reliability and validity” (Wertz, 2005: 176). He suggested phenomenology is particularly suited to counseling psychologists whose “work brings them close to the naturally occurring struggles and triumphs of a person” (Wertz, 2005: 176). Our findings contribute further to the formalization of justifiable norms in health and social science and aid phenomenology in moving toward Giorgi’s perspective that a method should be clear such that it is usable by any researcher. This is not to say that all phenomenological studies should look the same; rather, that researchers can turn to these findings as a source of greater scientific clarity, especially given psychologists’ concerns that qualitative samples are often “too small.”

Moreover, we agree with Wertz’s suggestion that phenomenology is well-suited to the work of researchers who
aim to be close to the struggles and successes of the individuals with whom we work. Psychologists concern themselves with individuals in contexts, and this sentiment is certainly no less true in other social sciences. For example, Brown et al. (2006) used phenomenology to better understand the difficult, life altering meaning of waiting for a liver transplant. These authors were able to use phenomenology to closely understand “waiting [for a liver transplant] as a time apart . . . a time to prepare for both life and death” (Brown et al., 2006: 132). Phenomenological methods represent an avenue of inquiry by which we center the experience of individuals and draw some shared conclusions through analysis. This method allows the voices of participants struggling with physical illness, health care access in marginalized communities, psychological distress after resettlement, and numerous other phenomena to be visible rather than reduced. Brown and colleagues accomplished this closeness to the experience of waiting for a liver transplant through phenomenology. However, researchers will benefit from work that clarifies how to justifiably carry out the method. The findings of this study are crucial for researchers who aim to be close to and contextualize the “struggles and triumphs” of individuals. Researchers can draw from our findings in conjunction with variant-specific guides as a tool to better communicate the value of their work to corners of social or other science that still undervalue insight drawn from phenomenology and qualitative methods.

**Recommendations for researchers**

We also want to turn attention to the meaning of these results. In part, they indicate that many researchers need further guidance for phenomenological methods. We offer the following suggestions for conducting phenomenological research:

- Researchers must familiarize themselves with the philosophical assumptions of phenomenology and conduct research in accordance with variant standards. The average quality score for these studies was 4.23 (with a high score of 12 in the sample). This seems concerning; however, researchers may develop greater aptitude with phenomenology through closer attention to the assumptions and procedures of the philosophy and method.
- We do not aim to make a decision about using focus groups; however, our data indicate that manuscripts with focus groups tend to be of less quality. This finding cannot be ignored. If researchers want to use focus groups in phenomenology, they too must pay closer attention to the steps and procedures of the method and invest effort in justifying the use of focus groups to access individuals’ expressions of consciousness.
- Recommending a concrete sample size is antithetical to this work and contrary to other theoretical positions on the matter (cf. Sim et al., 2018). Our aim was to determine what is typical, not set a benchmark for what a flexible method ought to look like in each instance. Samples must be sufficient to answer a research question. However, our findings can be considered alongside calls for small samples. We suggest that researchers can adequately conduct a phenomenological study with the average sample sizes (see Table 1) for each of the variants or with samples smaller than these averages. We caution researchers against going above this average as, overall, our data show a negative relationship between sample size and study quality. Researchers should be certain they are able to answer their research questions with the data they have, first and foremost. However, with phenomenology, it appears samples should be smaller.

**Conclusion**

Although our study is the first systematic review of samples in phenomenological research and connects sample characteristics to a quantified study quality variable, this work is not without limitations. We sampled studies only from PsycINFO, which indexes a variety of journals in psychology and related fields. We may have uncovered additional, different insights were another database of articles used. Our coding procedure also relied upon two individuals, with a third researcher to observe and guide the process. Although internal consistency and interrater reliability were appropriate, we cannot escape the reality that other coders may have derived different quality ratings. Similarly, our results contain a subset of phenomenology studies that were randomly sampled to allow adequate power. As a result, our data speak to phenomenological methods in broader strokes; thus, future researchers should endeavor to explicitly study samples and quality within specific variants of the method and types of data. This is especially important as some categories had a few representatives (e.g. transcendental and those studies using focus groups). Although the predominant use of random selection may suggest that a few researchers use transcendental phenomenology in manuscripts indexed in PsycINFO and that focus groups are rare, it is plausible our procedure excluded some of these manuscripts. In addition, we must remain cognizant of the diversity that is phenomenology. Numerous philosophers’ works fall within the scope of phenomenology, and because our empirical efforts focus on phenomenological methods, we do not directly attend to the breadth of phenomenologists (e.g. Merleau-Ponty, 1945). We coded primary variants with method-specific texts available; however, future efforts could adopt a more inclusive perspective on phenomenological research methods. Moreover, we found studies that predominantly utilized interview data; however, alternative data sources for the expression of consciousness (e.g. writings, photographs, etc.) may also be utilized in phenomenology. That we did not find many such studies in our sample is a limitation and
should not indicate a need to avoid using these data in phenomenology. Related to these latter points about the diverse reality of phenomenological studies, we also quantified quality. Although this facilitated our research purpose and offers meaningful insight into samples in phenomenological methods, it does represent a limitation of our approach. Finally, we used our explanatory, qualitative phase as a supporting component. Future research may use this approach to code more manuscripts and contribute further to understanding samples and quality in phenomenological research. Moreover, future research in this arena could attend to emergent technological realities in phenomenology such as the use of qualitative data analysis software. Others have addressed the use of such software in phenomenology (e.g. Goble et al., 2012; Sohn, 2017), and researchers could assess software use as a correlate of sample size and an additional predictor of study quality.

As such, our ratings of quality are not intended to impugn authors, editors, or editorial boards in instances of poorly rated studies. Rather, we see this as a point for highlighting needed growth, methodological rigor, and specificity. Conducting phenomenological research well requires adherence to the developed ways of implementing the method. Researchers have an obligation to attend to these ways of implementing a method while leaving room for creativity, and editors as well as reviewers must develop competence to adequately assess samples and study quality in phenomenological research. Despite the need for more research, this study serves as a reference for understanding what samples are common and acceptable for phenomenological methods used to study health, psychological, social, and related phenomena.

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

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

ORCID iD
Theodore T Bartholomew https://orcid.org/0000-0002-1979-5689

References
Blaikie N (2018) Confounding issues related to determining sample size in qualitative research. *International Journal of Social Research Methodology* 21: 635–641.
Bradburry-Jones C, Sambrook S and Irvine F (2009) The phenomenological focus group: An oxymoron? *Journal of Advanced Nursing* 64: 663–671.
Brown J, Sorrell JH, McClaren J, et al. (2006) Waiting for a liver transplant. *Qualitative Health Research* 16: 119–136.
Cohen J (1988) *Statistical Power for the Behavioral Sciences* (2nd edn). Routledge.
Colaizzi PF (1978) Psychological research as a phenomenologist views it. In: Valle RS and King M (eds) *Existential Phenomenological Alternatives for Psychology*. New York: Oxford University Press, pp. 48–71.
Conlon C, Timonen V, Elliot-O’Dare C, et al. (2020) Confused about theoretical sampling? Engaging theoretical sampling in diverse grounded theory studies. *Qualitative Health Research* 30: 947–959.
Creswell JW and Poth CN (2017) *Qualitative Inquiry and Research Design: Choosing among Five Approaches* (4th edn). Thousand Oaks, CA: SAGE.
Creswell JW and Plano Clark VL (2017) *Designing and Conducting Mixed Methods Research* (3rd edn). Thousand Oaks, CA: SAGE.
EBSCOhost (2018) How Is Relevance Ranking Determined in EBSCOhost? 1 October. Available at: https://help.ebsco.com/interfaces/EBSCOhost/EBSCOhost_FAQs/relevance_ranking_determined_in_EBSCOhost
Ehlers MM, Nielsen CV and Bjerrum MB (2018) Experiences of older adults after hip fracture: An integrative review. *Rehabilitation Nursing* 43: 255–266.
Elo S and Kyngas H (2007) The qualitative content analysis process. *Journal of Advanced Nursing* 62: 107–115.
Faul F, Erdfelder E, Lang A-C, et al. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39: 175–191.
Finlay L (2008) A dance between reduction and reflexivity: Explicating the “phenomenological psychological attitude.” *Journal of Phenomenological Psychology* 39: 1–32.
Finlay L (2012) Unfolding the phenomenological research process: Iterative stages of “seeing afresh.” *Journal of Humanistic Psychology* 53: 172–201.
Finlay L (2014) Engaging phenomenological analysis. *Qualitative Research in Psychology* 11: 121–141.
Freeman M and Vagle M (2013) Grafting the intentional relation of hermeneutics and phenomenology in linguisticality. *Qualitative Inquiry* 19: 725–735.
Giorgi A (2009) *The Descriptive Phenomenological Method in Psychology: A Modified Husserlian Approach*. Pittsburgh, PA: Duquesne University Press.
Giorgi A, Giorgi B and Morley J (2017) The descriptive phenomenological psychological method. In: Willig C and Stanton Rogers W (eds) *The SAGE Handbook of Qualitative Research in Psychology*. London: SAGE, pp. 176–192.
Goble E, Austin W, Larsen D, et al. (2012) Habits of the mind and the split-mind effect: When computer-assisted qualitative data analysis software is used in phenomenological research. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 13: 1709.
Hannes K, Lockwood C and Pearson A (2010) A comparative analysis of three online appraisal instruments’ ability to assess validity in qualitative research. *Qualitative Health Research* 20: 1736–1743.
Heidegger M (1962) *Being and Time* (trans. J Macquarrie and E Robinson). London: SCM Press.
Holloway I and Todres L (2003) The status of method: Flexibility, consistency and coherence. *Qualitative Research* 3: 345–357.
Husserl E (1980 [1911]) *The Phenomenology of Internal Time-Consciousness*. Bloomington, IN: Indiana University Press.
Kottak C (2006) *Mirror for Humanity*. New York: McGraw-Hill.

Liberati A, Altman DG, Tetzlaff J, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology* 62: 1–34.

Lockwood C, Munn Z and Porrit K (2015) Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence-Based Healthcare* 13: 178–187.

Malterud K, Siersma VD and Guassora AD (2016) Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research* 26: 1753–1760.

Merleau-Ponty M (1962 [1945]) *Phenomenology of Perception* (trans. C Smith). London: Routledge & Kegan Paul.

Moustakas C (1994) *Phenomenological Research Methods*. London: SAGE.

Nilsson I, Danbjerg DB, Aagaard H, et al. (2015) Parental experiences of early postnatal discharge: A meta-synthesis. *Midwifery* 31: 926–934.

Palmer M, Larkin M, de Visser R, et al. (2010) Developing an interpretive phenomenological approach to focus group data. *Qualitative Research in Psychology* 7: 99–121.

Polkinghorne DE (1989) Phenomenological research methods. In: Valle RS and Halling S (eds) *Existential-Phenomenological Perspectives in Psychology: Exploring the Breadth of Human Experience*. New York: Plenum Press, pp. 41–60.

Reiners GM (2012) Understanding the differences between Husserl’s (descriptive) and Heidegger’s (interpretive) phenomenological research. *Journal of Nursing and Care* 1: 1–3.

Sim J, Saunders B, Waterfield J, et al. (2018) Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology* 21: 619–634.

Smith JA (2011) Evaluating the contribution of interpretive phenomenological analysis. *Health Psychology Review* 5: 9–27.

Smith JA, Flowers P and Larkin M (2009) *Interpretive Phenomenological Analysis: Theory, Method and Research*. Thousand Oaks, CA: SAGE.

Sohn BK (2017) Phenomenology and qualitative data analysis software (QDAS): A careful reconciliation. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 18: 2688.

Todres I and Galvin KT (2008) Embodied interpretation: A novel way of re-presenting meanings in phenomenological research. *Qualitative Research* 8: 568–583.

Vagle MD (2018) *Crafting Phenomenological Research* (2nd edn). Routledge.

Vagle MD and Hofsess BA (2015) Entangling a post-reflexivity through post-intentional phenomenology. *Qualitative Inquiry* 22: 334–344.

van Manen M (1990) *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*. New York: The State University of New York.

van Manen M (2014) *Phenomenology of Practice: Meaning-Giving Methods in Phenomenological Research and Writing*. New York: Taylor & Francis.

Wertz FJ (2005) Phenomenological research methods for counseling psychology. *Journal of Counseling Psychology* 52: 167–177.

Woods M, Paulus T, Atkins DP, et al. (2016) Advancing qualitative research using qualitative data analysis (QDAS) software? Reviewing potential versus practice in published studies using ATLAS.ti and Nvivo, 1994-2013. *Social Science Computer Review* 34: 597–614.

Wride JM and Bannigan K (2018) “If you can’t help me, so help me god I will cut it off myself . . .” The experience of living with knee pain: A qualitative meta-synthesis. *Physiotherapy* 104: 299–310.

**Author biographies**

**Eileen E Joy**, MA, is a doctoral candidate in counseling psychology in the Department of Educational Studies at Purdue University. Her research examines the roles of cultural context and poverty-related stressors in therapeutic processes and outcomes.

**Ellice Kang**, MEd, is a doctoral candidate in counseling psychology in the Department of Educational Studies at Purdue University. Her research interests include cultural and racial factors in mental health and identity development and social justice issues among marginalized populations.

**Jill Brown**, PhD, is a Professor of Psychological Science at Creighton University in Omaha, NE. Her teaching and research focus on the psychological underpinnings of culture and human development, specifically, kinship, adoption and socially distributed child care, relative deprivation and psychological health, and cognition and thinking across cultures. She is the director of African Studies and Past President of the Society of Cross-Cultural Research.