Data Saturation and Its Determination in Qualitative Research

Authors: Yang Liping, Qi Lidong, Zhang Bo, Yang Liping

Date: 2021-10-24T00:00:00+00:00

Abstract

In qualitative research, data saturation is commonly employed to assess the sufficiency of research data. However, in research practice, data saturation exhibits conceptual ambiguity and operational challenges. Serving as an indicator that the sample size extracted for a qualitative study has satisfied research requirements, data saturation can be categorized into four primary forms according to the chronological sequence of determination points in the research process: data saturation, coding or thematic saturation, meaning saturation, and theoretical saturation. Each of these four forms of data saturation possesses its distinct connotations, evaluation approaches, and determination criteria. The study contends that the sample size criterion for achieving data saturation should not be uniformly prescribed; its verification must be embedded within the specific research process; data saturation entails logical uncertainty, and moderate additional sampling facilitates further confirmation; and as a crucial indicator for evaluating research quality, data saturation is not applicable to all qualitative research.

Full Text

Data Saturation and Its Determination in Qualitative Research

YANG Liping, QI Lidong, ZHANG Bo
(School of Psychology, Nanjing Normal University, Nanjing 210024)

Abstract

In qualitative research, data saturation is commonly used to assess the adequacy of research data. However, in research practice, data saturation suffers from conceptual ambiguity and operational challenges. As an indicator that the sample

size extracted for a qualitative study has met research requirements, data saturation can be distinguished into four main forms according to the temporal sequence of determination during the research process: data saturation, code or thematic saturation, meaning saturation, and theoretical saturation. Each form of data saturation has its specific connotations, evaluation methods, and judgment criteria. This study argues that the sample size standard for achieving saturation should not be uniformly predetermined; its assessment must be embedded within the specific research process. Data saturation entails logical uncertainty, and moderate additional sampling helps to further confirm it. Although data saturation serves as an important indicator for evaluating research quality, it is not applicable to all qualitative research.

Keywords: qualitative research, data saturation, evaluation methods, judgment criteria

1 Introduction

Since the establishment of the Professional Committee for Qualitative Research in Psychology of the Chinese Psychological Society in January 2018, papers employing qualitative and mixed-methods research have increased substantially. In qualitative research, as in quantitative research, adequate sampling is a fundamental guarantee of research validity (Curtis et al., 2000). Insufficient sample size affects research quality, while excessive sample size wastes research resources. Data saturation serves as the marker of appropriate sample size in qualitative research. Premature termination of data collection before reaching saturation results in findings that lack necessary insight, reducing them to simple presentations of raw data (Sudaby, 2006/2016). Achieving data saturation means that, based on currently collected and analyzed data, further data collection will not help researchers develop deeper understanding of the narrative or theory, making continued data collection unnecessary (Corbin & Strauss, 2014). It should be noted that, unlike quantitative research, data collection and analysis in qualitative research are not two distinct phases but rather exhibit a cyclical, iterative characteristic. Data saturation in qualitative research is typically a judgment made based on analysis of previously collected data. Once data collection reaches saturation, data analysis also concludes.

Previously, sample size in qualitative research has been determined through two main approaches: empirical rules or saturation testing (Marshall et al., 2013). Based on empirical rules, existing studies have summarized appropriate sample sizes for different types of qualitative research. However, such literature mostly lacks clear description of underlying principles, and recommended sample size standards often vary considerably. For example, among researchers' recommendations, grounded theory studies suggest interview sample sizes ranging from 5 to 35, while case studies range from 4 to 30 (Sim et al., 2018). Some researchers argue that, due to the characteristics of qualitative research and differences between studies, establishing uniform sample size standards lacks theoretical and practical basis (Guest et al., 2006; Marshall et al., 2013).

Using data saturation as a methodological principle for stopping data collection and analysis in qualitative research has been widely accepted and employed (Saunders et al., 2018). Data saturation has even been called the "gold standard" and "action guide" for determining qualitative research sample size (Guest et al., 2006), and a "guarantee" of qualitative research quality (Morse, 2015). Some researchers propose that in qualitative research, the most common criterion for evaluating sampling adequacy is data saturation (Fusch & Ness, 2015; Morse, 2015). Simultaneously, data saturation is an important criterion for evaluating the quality of qualitative research papers (Hennink et al., 2017), and its appropriate use can improve qualitative research quality, such as rigor and validity (Parker, 2013; Sim et al., 2018). The American Psychological Association Publications and Communications Board Task Force on Journal Article Reporting Standards for Qualitative Research also noted that researchers should explain the basis for determining sample size and recommended using data saturation for testing (Levitt et al., 2018).

Although the concept and function of data saturation are widely recognized and frequently mentioned by qualitative researchers, in actual practice this concept and its criteria remain ambiguous (Hennink et al., 2017) and suffer from long-standing conceptual and operationalization problems (Saunders et al., 2018). On one hand, the concept of data saturation originated from grounded theory, initially appearing as theoretical saturation (Glaser & Strauss, 1967). However, as data saturation has been promoted as a criterion for sample size judgment in qualitative research, multiple indicators have developed, such as data saturation, code saturation, thematic saturation, and meaning saturation. Due to the diverse forms and standards of saturation, with complex and ambiguous relationships between different forms and standards, few researchers can systematically and thoroughly understand and master the concept of data saturation, often causing confusion in its application (Parker, 2013). On the other hand, in numerous research reports, researchers claim to have used data saturation as the criterion for sample size and to have achieved it, yet rarely specify which type of saturation was reached, its judgment method, or its justification (Constantinou et al., 2017; Morse, 2015). Consequently, some researchers propose that transparency in reporting data saturation in qualitative research should be increased. When authors claim to have achieved saturation, they should provide more detailed descriptions clarifying the form of saturation, determination methods, and basis, enabling readers to evaluate and further verify the claims. This is also a basic requirement for normative awareness in qualitative research (Fusch & Ness, 2015; Hennink et al., 2017; Kerr et al., 2010; Malterud et al., 2016).

However, since previous studies lack operational descriptions and practical guidance for the evaluation and reporting of data saturation (Carlsen & Glenton, 2011; Hennink et al., 2019), requiring researchers to accomplish this is unrealistic. Given the importance of data saturation to qualitative research and its current operational status, it is necessary to explore and explain the definitions, criteria, and reporting methods of various forms of data saturation to

help researchers better understand and use the concept, provide evidence-based criteria for sample size judgment in qualitative research, and thereby improve the standardization and quality of qualitative research in psychology.

Based on a review of literature on data saturation and the author team' s long-term experience conducting qualitative research, this study attempts to clarify the concept of data saturation in qualitative research, differentiate its various forms, explore their relationships through comparison, and investigate methods for determining and reporting data saturation, aiming to provide reference for saturation judgment in qualitative research.

2 Four Forms of Data Saturation in Qualitative Research

Data saturation is a sign that the sample size extracted for a qualitative study has met research needs. Grounded theory first raised the issue of data saturation, using theoretical saturation as its criterion (Glaser & Strauss, 1967). However, with the development of qualitative research, the concept of data saturation is no longer limited to theoretical saturation proposed by grounded theory but has further evolved to include multiple forms such as data saturation, code or thematic saturation, and meaning saturation. Different forms of data saturation vary in their meanings, orientations, analytical methods used, and judgment criteria (Saunders et al., 2018).

2.1 Theoretical Saturation

In qualitative research, grounded theory first proposed the issue of research data saturation. Grounded theory is a research method that seeks core concepts reflecting social phenomena through systematic data collection and establishes theory by forming connections among these concepts (Chen Xiangming, 2000). The primary purpose of grounded theory research is theory construction, requiring that the constructed theory achieve consistency and coordination both internally and with external related theories. Based on this, grounded theory founders Glaser and Strauss (1967) first proposed the concept of theoretical saturation, explaining it as "if sampling continues, no new categories or related themes will emerge."

The concept of theoretical saturation in grounded theory means that during data collection, no new themes or insights can be developed, all properties of relevant conceptual categories and their relationships have been explored, exhausted, and determined, continued data acquisition cannot reveal new properties or provide deeper understanding of the emerging theory, and the new theory has become comprehensive and credible (Hennink et al., 2017; Morse, 2015). Hennink et al. (2019) point out that theoretical saturation depends on sample adequacy, requiring researchers to discover sufficient, rich, logical, and meaningful data to support the emerging theory. Saunders et al. (2018) argue that theoretical saturation is achieved when all concepts representing the theory are fully reflected in the data.

The judgment of theoretical saturation in grounded theory is rooted in the cyclical, spiral evolution process of qualitative research and is closely related to grounded theory' s requirement for theoretical sampling (Saunders et al., 2018). In grounded theory research, after data collection reaches a certain extent and undergoes coding, comparison, categorization, and re-comparison, the outline of the emerging theory is initially sketched to guide further theoretical sampling. Then, based on newly collected data, the emerging theory is tested, improved, and refined. This cyclical evolution gradually removes weak links in the emerging theory. It is a continuous process of comparing and validating the emerging theory against original data, continuously optimizing and perfecting it (Chen Xiangming, 2000). When the emerging theory can basically explain all data and researchers can no longer discover new concepts representing the theory from the data but only repetitions of existing concepts or properties, theoretical saturation is considered achieved and sampling can stop.

Some researchers point out that in grounded theory research, driven by theoretical sampling, conceptual categories or emerging theories evolve cyclically and continuously improve. When relationships between conceptual categories and their meanings gradually become clear, theoretical saturation is reached (Morse, 2015). This shows that theoretical saturation is a saturation judgment based on theoretical completeness, made at a later stage of data collection and analysis with a high level of theoretical generalization (Saunders et al., 2018).

2.2 Data Saturation

Data saturation refers to the point in the data collection process where no new data are generated and information redundancy gradually appears. For example, in interviews, researchers repeatedly hear the same narratives and thus judge that data saturation has been reached, stopping data collection (Grady, 1998; Jackson et al., 2015). Data saturation focuses on whether new information will continue to emerge and represents a judgment of saturation made during the data collection stage. Data collection occurs at a relatively early stage of the entire research process, and researchers make judgments based on the assumption that "some themes may emerge during subsequent data analysis." However, such judgments are often inaccurate because during data analysis, various codes, code properties, and relationships between codes will continue to change, including merging of codes, expansion of meanings, and exclusion of meanings (Saunders et al., 2018). Therefore, using data saturation as a criterion for data saturation judgment has obvious drawbacks.

2.3 Code or Thematic Saturation

Code or thematic saturation, sometimes called category saturation, indicates that during data analysis, the scope of codes or themes has been basically determined. Only content repetitive of existing codes or themes can be extracted from newly collected data, no new codes or themes emerge, and the codebook has become relatively stable (Urquhart, 2012; Hennink et al., 2017). Code or

thematic saturation focuses on the quantitative repetitiveness of codes or themes generated during data analysis, with the criterion being that no new codes or themes appear during data analysis.

Code or thematic saturation represents progress in reliability compared to data saturation. Coding already involves analysis of data, while themes are results that emerge after coding reaches a certain level. Compared to data saturation, which judges solely based on repetitiveness of data or materials themselves, code or thematic saturation has incorporated a certain level of analysis. However, using code or theme repetition as the basis for saturation judgment still has problems and can easily lead to false saturation, resulting in insufficient information or materials to construct a sound theory. This is because code or thematic saturation only provides a general outline of the research question and can often be achieved through relatively few interviews (Saunders et al., 2018). Therefore, achieving only code or thematic saturation is insufficient; more data are needed for researchers to fully understand the depth, richness, and complexity of the problem (Emmel, 2015; Hennink et al., 2017).

Like data saturation, code or thematic saturation only assesses saturation based on the quantity or frequency of codes or themes—that is, saturation judgment based solely on the breadth and scope of collected materials—lacking evaluation of the meaning of codes or themes. When a theme first emerges from the data, researchers often may not achieve deep understanding of that theme and need to further collect and analyze data to discover richer and deeper knowledge about the theme or conceptual category (Hennink et al., 2017; Kerr et al., 2010).

2.4 Meaning Saturation

Meaning saturation means that during data collection and analysis, researchers have fully understood the series of codes or themes developed, and no new information appears regarding the meanings of codes or themes and their relationships (Hennink et al., 2017). Hennink et al. (2017) believe that achieving meaning saturation requires a cyclical, evolutionary process of sampling, data collection, and data analysis, continuously monitoring the diversity, clarity, and depth of data, emphasizing data collection targeting currently poorly understood information, codes, or themes. Meaning saturation is a data saturation judgment based on the completeness of code or theme meanings during data collection and analysis, focusing on deep understanding of data. Using this indicator to judge saturation facilitates discovering and presenting the complete meaning of data.

Hennink et al. (2017) compared code saturation and meaning saturation. In 25 in-depth interviews, code saturation was reached at the 9th interview, meaning all codes were identified and the scope of main thematic issues was determined. However, 16-24 interviews were needed to achieve meaning saturation, where researchers truly understood the meaning of codes and had rich understanding of the research question. Hennink et al. believe that code saturation only

represents "having heard it all," while achieving meaning saturation requires "understanding it all." This again demonstrates that in saturation testing, relying solely on "some codes or themes begin to repeat, no other new codes or themes can be discovered (Kerr et al., 2010)" as a judgment criterion is insufficient. Researchers should also determine whether the definition and content of each code or theme have been adequately explored and understood—that is, "the explanatory dimensions or understandings of each code or theme begin to repeat, no other interpretations of the code or theme are discovered, and the code or theme thus reaches meaning saturation."

The above four forms of data saturation address information redundancy or theoretical connotation in qualitative research data collection and analysis processes (Sim et al., 2018), appearing at different stages of the research process (Saunders et al., 2018). Data saturation directly focuses on the data collection process; code or thematic saturation is based on preliminary data analysis; meaning saturation further emphasizes depth of data analysis based on data collection and analysis (which often occur simultaneously or iteratively in qualitative research); and theoretical saturation focuses on higher-level theory construction, emphasizing adequacy in discovering conceptual categories and their properties (theoretical meanings) during theory construction.

Different forms of data saturation have different judgment foci. Data saturation, code or thematic saturation are judged based on repetition or frequency of data, codes, or themes; meaning saturation is judged based on depth of meaning of codes or themes; theoretical saturation is judged based on completeness of properties of concepts in the emerging theory, self-consistency of the emerging theory, and consistency between the emerging theory and previous theories. Overall, data saturation and code or thematic saturation focus on breadth of collected data, while meaning saturation and theoretical saturation focus on depth of research data (Saunders et al., 2018). Meaning saturation is closer to theoretical saturation, but they have essential differences. Meaning saturation focuses on depth of meaning of codes or themes, while theoretical saturation further focuses on completeness, predictive power, and explanatory power of the emerging theory developed from codes or themes. Moreover, the concept of theoretical saturation originates from and mainly applies to grounded theory, whereas meaning saturation is not limited to the grounded theory context and applies to judging data saturation in more types of qualitative research (Hennink et al., 2017).

3 Determination of Data Saturation

3.1 Determination of Theoretical Saturation

As mentioned above, the concept of theoretical saturation originates from grounded theory. In grounded theory research, theory development and refinement must be based on systematic data collection and analysis, achieved through the interaction of data collection and data analysis. The main

analytical method is repeated comparison between data and data, theory and theory, extracting conceptual categories and their properties based on relevance between data and theory, also known as the "constant comparative method" (Chen Xiangming, 2000). Constant comparison is the main characteristic of grounded theory research, and theoretical saturation depends on this ongoing comparative process (Bowen, 2008).

After a series of comparisons, researchers outline a new theory and continuously compare it with early collected data and newly collected data to refine the theory. When researchers find that the theory can explain most (or all) of the original or new data, they can judge that the study has reached theoretical saturation.

Bowen (2008), using a grounded theory study as an example, conducted data collection, coding analysis, theoretical sampling, constant comparison, and other operations according to grounded theory research requirements, and detailed the criteria for determining theoretical saturation. He argued that four conditions must be met to determine that research data have reached theoretical saturation: 1) relevant conceptual categories are reflected in more than 70% of interviews; 2) respondents identify with the research results and provide good feedback; 3) the results align with previous research findings; 4) new respondents begin to repeat content similar to previous respondents. Bowen also recommended that if a standard such as 70% is used as a criterion for saturation, it is best combined with other forms of saturation criteria.

3.2 Determination of Data Saturation

Researchers predict data saturation based on the content narrated by interviewees during data collection and 预设 of possible themes that may emerge during data analysis. When researchers repeatedly hear the same narratives, they begin to consider whether to judge that the research has reached data saturation (Jackson et al., 2015). As mentioned earlier, researchers' early understandings may change dramatically during data analysis, making it unreliable to judge research saturation based solely on these characteristics. However, applying data saturation to assess saturation of single interviews or qualitative research analyzing individual life histories (such as individual psychobiographies) seems feasible.

3.3 Determination of Code or Thematic Saturation

Code or thematic saturation is a form of data saturation frequently used in qualitative research, with related criteria studied and···

3.3.1 Using Empirical Research Results as Criteria Some studies have used retrospective empirical analysis to systematically present the evaluation, reporting, and verification processes of data saturation, summarizing the sample sizes needed to achieve code or thematic saturation. For example, Guest

et al. (2006) conducted 60 in-depth interviews, and after interviews were completed, judged saturation in groups of 6 interviews based on the development of themes and important themes (with important themes determined by the number of codes they contained). Results showed that using thematic analysis, basic elements of themes were already presented after the first 6 interviews. After 12 interviews, 88% of themes and 97% of important themes were identified, the codebook structure became stable and rarely needed modification, and the study reached thematic saturation. Guest et al. (2017) used a similar method to analyze data from 40 focus group interviews, finding that 84% of themes could be identified after 3 focus groups, and 90% of themes after 6 focus groups.

Francis et al. (2010) explored saturation of conceptual categories in theory-based interview studies, finding that overall category saturation was achieved after 17 in-depth interviews. Coenen et al. (2012) used maximum variation sampling to verify the sample size needed to reach saturation in focus group interviews, finding that coding saturation was achieved after 5 focus groups. Hennink et al. (2017, 2019) conducted retrospective analyses of sample sizes needed for semi-structured in-depth interviews and focus group studies, finding that among 25 in-depth interviews, coding saturation was reached at the 9th interview; while among 10 focus group interviews, saturation was reached at the 4th interview. Similarly, researchers found that in relatively homogeneous groups, sufficient common themes could be obtained through 16 samples, while 20-40 samples were needed in cross-cultural contexts (Hagaman & Wutich, 2017); in exploring more abstract concepts, 92% of codes could be summarized through interviews with 12 samples (Ando et al., 2014).

Most of the above studies used retrospective analysis, using all codes or themes obtained in the research as the base, calculating the proportion of codes or themes obtained after a certain interview to the total number to judge data saturation, aiming to provide standard references for subsequent research based on results from multiple empirical studies and researchers' long-term accumulated experience. However, in actual operation, saturation judgment is process-based —that is, saturation must be assessed during the cyclical process of data collection and analysis to determine when to stop sampling. Retrospective analysis is only equivalent to "hindsight" or "second-guessing." Although such studies have yielded numerous empirical experiences about sample size, as many researchers emphasize, their findings cannot be directly applied as generalizable sampling requirements to other qualitative research because data saturation is influenced by many factors including research questions, research purposes, sample characteristics, sampling homogeneity, interview methods, coding characteristics, researchers, and data features. Sample size should be carefully determined according to the research' s methodological characteristics, epistemological stance, and research resources (Guest et al., 2006; Hennink et al., 2017).

3.3.2 Using Whether New Codes or Themes Emerge as Criteria Some researchers tend to pursue quantitative indicators, judging data saturation by

comparing newly obtained information with already obtained information—that is, when the proportion of new codes emerging in an interview to the total number of determined codes reaches a certain standard, the research can be judged to have reached data saturation. Guest et al. (2020) found that in qualitative research data collection, most new information appears early, typically following a diminishing returns curve in the short term. After a certain amount of data collection or analysis, the amount of new information declines sharply. Therefore, they selected the first 4 (or 5, 6) data collections as the denominator, used each subsequent 2 (or 3) interviews as the numerator, and used 5% or 0% as threshold values to judge data saturation. When the amount of new information obtained in 2 new interviews accounted for less than 5% of the total information obtained in the first 4 interviews, the research was judged to have reached code or thematic saturation. The selection of saturation threshold values (5% or 0%) referenced the setting of the p-value in quantitative research. Guest et al. proposed that the choice of specific standards is autonomous, and qualitative researchers can determine them according to research realities.

Judging data saturation by calculating the ratio between newly emerged information and already obtained information is relatively easy to operate. Researchers can also use charts to assist in making research-purpose-appropriate judgments about saturation based on the emergence of new codes or themes during the research process (Guest et al., 2020; Hennink et al., 2017, 2019). For example, Hennink et al. used bar charts to represent saturation of codes or themes in terms of breadth, with the number of new codes as the vertical axis and interview order as the horizontal axis. The height of bars represented the number of new codes or themes obtained in each interview (Hennink et al., 2017). [Figure 1: see original paper] shows the development process of codes or themes in that study as interviews continued. It can be seen that the first interview obtained a large number of codes, and subsequent interviews clearly showed diminishing returns. Starting from the 16th interview, cases of providing no new information appeared consecutively, indicating that the scope of codes or themes had been basically determined and the study could be judged to have reached code or thematic saturation. If interviews continued, although the possibility of new information emerging is not completely ruled out, qualitative research does not pursue exhaustive acquisition of research data, only relatively sufficient data to develop and verify concepts and theories.

Constantinou et al. (2017) used the Comparative Method for Themes Saturation (CoMeTS) to judge thematic saturation. First, they compared analysis results from each new round of interview data with previous analysis results, identifying the number of new themes and repeated themes obtained in the most recent interview. Saturation was considered achieved when no new themes emerged. Then, researchers scrambled and reordered the interview data, conducted analysis again, and tested thematic saturation. Results showed that when analyzing data in natural interview order, thematic saturation was reached at the 5th interview, but after multiple random reorderings of interview data, saturation was reached at the 7th or 8th interview. The results indicate that the order

of interview data analysis affects saturation judgment. It should be noted that interview data collection and analysis are typically conducted concurrently, and scrambling interview data order does not conform to qualitative research logic. However, incorporating comparative methods in the determination process does help further confirm data saturation.

3.3.3 Using Saturation Coefficients as Criteria Some researchers have attempted to explore data saturation in qualitative research using more complex calculation formulas. Viet-Thi et al. (2017) used Monte Carlo simulation (also called random simulation method) in open surveys to predict the number of themes new participants could provide. Lowe et al. (2018) measured thematic saturation in qualitative research based on mathematical statistical models, providing measurement methods for saturation between datasets, within datasets, and during research projects. Namey et al. (2016) demonstrated a method for calculating sample sizes needed to reach thematic saturation by using bootstrap simulation to randomly generate 10,000 samples from each qualitative research dataset. Other researchers have proposed sample size evaluation methods based on binomial distribution (Fugard & Potts, 2015; Galvin, 2015).

The above calculation methods for data saturation all depend on probability theory and random sampling (Fugard & Potts, 2015; Galvin, 2015; Lowe et al., 2018). However, because qualitative research uses non-probability sampling (Guest et al., 2020) and has characteristics such as openness, it is not suitable for probability theory or statistical inference (Blaikie, 2018; Sim et al., 2018). Additionally, due to limitations of research traditions and knowledge scope, complex mathematical statistical analysis models are neither convenient for humanities and social science researchers to operate and use, nor conducive for corresponding readerships to accept and understand.

Some researchers have attempted to introduce the Jaccard coefficient as an indicator for judging data saturation (Liu Tianfang & Yang Liping, 2018). The Jaccard coefficient originally refers to the ratio of the intersection to the union of two given datasets A and B, with larger ratios representing higher similarity between the two sets (Zhang Meng & Li Lingjuan, 2018). The Jaccard coefficient calculation formula is:

$$J(A,B) = |A \cap B| / |A \cup B| = |A \cap B| / (|A| + |B| - |A \cap B|)$$

In data analysis, the Jaccard coefficient can be used to calculate code similarity of data obtained from different interview rounds. The greater the similarity between the newly obtained dataset and previously collected data, the higher the Jaccard coefficient, indicating higher saturation of data coding. This operational approach aligns with the concept of coding saturation. In analyzing English materials, English words are mainly used as analysis units to calculate the Jaccard coefficient, while Chinese uses individual characters or codes as identification units. Currently in practice, the Jaccard coefficient is mostly only

a reported value. As for what value of the indicator can be recognized as coding saturation, there is no unified standard and further exploration is needed. Additionally, using the Jaccard coefficient as a criterion has other problems. Given the continuous and exploratory nature of qualitative research processes, using the union of the newly obtained dataset B and all previously obtained data A as the denominator makes the denominator too large, causing the Jaccard coefficient to be too small and thus losing sensitivity to changes in data saturation during the research process.

To compensate for the shortcomings of the Jaccard coefficient, researchers have constructed another saturation indicator that is simpler to operate and more sensitive to changes in data saturation, called the S (Saturation) coefficient, to judge coding saturation in qualitative research. For two given sets A and B, where B represents the newly obtained dataset and A represents all data collected before B, the S coefficient is the ratio of the intersection of the two datasets' codes to dataset B. Larger ratios represent higher repetition of codes in dataset B with previous sets. The S coefficient calculation formula is:

$$S(A, B) = |A \cap B|/|B|$$

Data saturation as a process-based evaluation indicator will continuously increase or accumulate as the research process progresses. The S coefficient can well reflect the trend of data saturation gradually increasing as sample size grows. During research, saturation can be recalculated after each additional sampling, thereby monitoring the contribution of new samples to code or theme development. In interviews with homogeneous samples, coding saturation can be recognized when the S coefficient reaches 95% or above for 3 consecutive individual interviews (or 2 consecutive focus groups). The 95% here references the setting of the p-value in quantitative research, while requiring "3 consecutive times" or "2 consecutive times" follows the principle of "appropriate additional sampling." However, this remains an empirical judgment that awaits verification and support from more studies.

3.4 Determination of Meaning Saturation

Hennink et al. (2017, 2019) used retrospective empirical analysis to explore the sample sizes needed to achieve meaning saturation in in-depth interview and focus group studies, comparing them with coding saturation. Results showed that 16-24 individual interviews were needed to achieve meaning saturation, and meaning saturation was barely reached after 10 focus group interviews. Typically, the sample size needed to achieve meaning saturation is much higher than that for coding saturation.

Researchers use tables to record the development and changes of meaning units (secondary codes) during the cyclical process of data collection and analysis, called meaning saturation grids (Brod et al., 2009; Hennink et al., 2017, 2019).

As shown in , the first column lists developed meaning units, and subsequent columns show new connotations added to meaning units (or different categories) by data collected at stages 1-6, 7-9, 10-12, and after 12.

Table 1 Meaning Saturation Grid (Hennink et al., 2017)

Meaning Unit	Meaning Unit Meanings	Interview 1-6	Interview 7-9	Interview 10-12	After Interview 12
Feeling good	No disease (1)	Feeling good (3)	Viral load stable (3)	Having disease so visiting doctor (3)	Having drug supply (4)
Time	Travel/stay time (1)	Taking leave (9)	Consultation time (1)	Time for other things (1)	Clinic scheduling (1)
	Time for other treatments (2)	Waiting time (3)	Lab test time (3)	Wasting time (3)	
Work commitment	Long working hours (5)	Employer questioning leave (13)	Work more important (6)	Needing to work during appointment time (6)	Inflexible work scheduling (6)
	Work leave regulations (6)	Avoiding work trouble (6)			
Disclosure stigma	Disclosure (1)	Clinical confidentiality (7)	Workplace disclosure (11)	Negative disclosure experiences (13)	Selective disclosure (1)
	Disclosure to other patients (8)	Public exposure (17)	Disclosure to relatives and partners (1)	Disclosure to friends (9)	Hiding HIV due to stigma (1)
	Fear of violence from disclosure (3)	Delaying disclosure (3)	Other disclosure situations (6)		

Meaning saturation grids can be established during the cyclical process of data collection and analysis. Developed meaning units are filled into the first column one by one, and sub-codes or free nodes of newly discovered meaning units are filled into corresponding positions in the grid according to their emergence order. If no new codes appear in an interview, the corresponding cell is left blank. The sign that a single meaning unit has reached saturation is that no new sub-codes or free nodes for that meaning unit appear in subsequent interviews. When all (or a certain proportion of) cells appear blank, representing that no new information will be added as interviews continue, the research data can be judged to have reached meaning saturation. In Table 1, the properties of the second meaning unit "time" were basically determined by the 9th interview, with no further additions afterward, indicating that this meaning unit has reached saturation. It should be reminded that in qualitative research processes, codes are always in a state of continuous adjustment, which can be reflected in meaning saturation grids.

Meaning saturation grids record the development process of meaning units and their meanings, helping researchers judge whether specific meaning units have reached saturation. However, such grids are mostly only kept in researchers' memos. Due to excessive content and complexity, they are difficult to present completely in journal articles. Hennink et al. (2017) therefore created another more referential way to present meaning saturation, shown in [Figure 2: see original paper]. The horizontal axis represents interview order, the vertical axis presents different meaning units, hollow circles indicate where a meaning unit was officially created, and solid circles indicate where a meaning unit reached meaning saturation—that is, where new information about that meaning unit no longer appeared in subsequent interviews, at which point the relevant information of that meaning unit can be considered to have reached saturation.

4 Discussion

Emphasizing data saturation in qualitative research has three significance aspects. First, it evaluates the adequacy of research data to guide the sampling process. Second, it helps researchers judge whether their understanding of research questions and related concepts is deep, reflecting research quality. Third, it helps researchers report the adequacy of research sampling in papers, accepting readers'scrutiny and facilitating follow-up research by subsequent researchers. However, in actual practice, different forms of data saturation are often confused, and problems such as inappropriate evaluation methods or vague operations are widespread.

4.1 Sample Size Standards for Achieving Data Saturation Should Not Be Uniformly Predetermined; Assessment Must Be Embedded in the Specific Research Process

Regarding saturation determination, some empirical research results or statistical methods attempt to provide sample size standards for achieving data sat-

uration before research begins. However, as mentioned above, data saturation in qualitative research is influenced by many factors, and each study has its individuality or particularity. For example, in interview studies, themes are closely connected to interviewees; interviewees' identities and contexts help 赋予 themes meaning and importance. Themes are "attributes" of interviewees, and the two correspond to each other (Byrne, 2015; Hammersley, 2015; Sim et al., 2018). During data analysis, themes are not static; theme naming, connotation, extension, and theoretical contribution to research questions are continuously developing, and the importance of specific themes and their relevance to research questions also change (Hammersley, 2015). Therefore, the depth of exploration of research questions and sample size do not follow a simple linear relationship, and the number of interviewees cannot alone serve as a basis for judging data saturation (Sim et al., 2018). The sample size to achieve research data saturation cannot be determined before research begins (Braun & Clarke, 2021).

In specific research processes, researchers should consider multiple forms of data saturation based on their study' s characteristics. For example, use code or thematic saturation to assess the breadth and scope of research data, and use meaning saturation to assess the depth of exploration of codes or themes. If the research aims to construct theory, theoretical saturation should also be tested through constant comparison. Embedding assessment within the research process and comprehensively using multiple forms of data saturation also has precedents in foreign studies, such as combining data saturation with theoretical saturation (Goulding, 2005; Morse, 2015).

4.2 Data Saturation in Qualitative Research Entails Logical Uncertainty; Appropriate Additional Sampling Helps Further Confirmation

Data saturation is a prediction about future interview data acquisition based on already obtained data. The "logical uncertainty" here means that predicting the necessity of continued data collection and analysis based on current data collection and analysis depends to some extent on researchers' subjective judgment, and its accuracy cannot be further proven (Saunders et al., 2018). Although researchers can clearly record and present the data collection and analysis process to provide evidence for research data saturation, data saturation in qualitative research is hardly an absolutely accurate judgment.

Because data saturation entails logical uncertainty, some researchers believe that additional sampling should be appropriately conducted after data have reached or basically reached saturation to verify saturation. For example, Jassim and Whitford (2014) found that research data had reached thematic saturation after 10 interviews but continued with 2 additional interviews to confirm saturation. Similar approaches have been used by other researchers (Bragaru et al., 2013; Jackson et al., 2000; Vandecasteele et al., 2015). However, Saunders et al. (2018) point out that oversampling practices also have problems to some extent, potentially causing conceptual ambiguity about data saturation. Nevertheless, appropriate additional sampling remains an effective strategy for addressing

logical uncertainty in data saturation. Typically, after making a judgment that research data have reached or basically reached saturation, researchers can continue with 2-3 additional individual interviews or 1-2 additional focus groups with homogeneous samples to further confirm data saturation.

4.3 As an Important Indicator for Evaluating Research Quality, Data Saturation Is Not Applicable to All Qualitative Research

Qualitative research is an "umbrella concept" containing various research methods, such as grounded theory, phenomenological research, discourse analysis, thematic analysis, ethnography, narrative research, focus groups, life history, and psychobiography (He Wuming & Zheng Jianhong, 2019). Some researchers believe that the concept of data saturation is not applicable to narrative analysis, interpretative phenomenological analysis, etc. (Saunders et al., 2018; van Manen et al., 2016). Thematic analysis, grounded theory, and focus groups use inductive thinking, collecting information from numerous respondents and focusing on adequacy of theory development, thus better fitting the operational definition of data saturation. Psychobiography and narrative analysis focus on individuals, paying more attention to the completeness of individual stories, which current data saturation concepts do not seem to include (Saunders et al., 2018). As for interpretative phenomenological analysis, although it also uses methods such as extracting themes and clarifying relationships between themes for data analysis, it emphasizes obtaining complete and rich individual understandings of life experiences, has an idiographic orientation, and stresses detailed examination of each case (Hou Liqi et al., 2019). Whether this type of research is suitable for the concept of data saturation requires further exploration and discussion. Therefore, some researchers worry that applying the concept of data saturation indiscriminately to all qualitative research may cause it to lose its consistent utility (Saunders et al., 2018). Future attention needs to be paid to examining, judging, and testing data adequacy in different types of qualitative research.

5 Conclusion

Qualitative research is not merely a new research method added within the framework of traditional empirical research but represents an entirely new methodology in psychology. It demonstrates different characteristics from empirical research in every aspect: research topic selection, literature review, establishment of research relationships, data collection, data analysis, results and discussion, validity testing, and research reporting. Data saturation and its testing are just one research 环节. Only by conducting in-depth and meticulous research on each 环节 in the qualitative research process can the standardization of qualitative research operations be improved and research quality enhanced.

References

Chen, X. M. (2000). *Qualitative research methods and social science research.* Beijing: Educational Science Publishing House.

Hou, L. Q., Tang, X. F., He, L., & Jia, X. M. (2019). The application of interpretative phenomenological analysis in China: Systematic review and guidelines. *Advances in Psychological Science*, (11), 1826.

He, W. M., & Zheng, J. H. (2019). Qualitative research in psychology: History, current status, and prospects. *Psychological Science*, 4(04), 1017.

Liu, T. F., & Yang, L. P. (2018). Aging psychology constructed based on age: A report of an in-depth interview study. *Psychological Science*, (05), 1207.

Sudaby, R. (2006). What grounded theory is not (Zeng Xianju & Han Wei, Trans.). In X. Y. Xu, B. Ren, & L. Lü (Eds.), *Management theory construction* (pp. 145-162). Beijing: Peking University Press.

Zhang, M., & Li, L. J. (2018). Community division algorithm based on improved Jaccard similarity coefficient matrix. *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)* (06), 96.

Ando, H., Cousins, R., & Young, C. (2014). Achieving saturation in thematic analysis: Development and refinement of a codebook. *C*, 1.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Q*, 137.

Byrne, D. (2015). Response to Fugard and Potts: Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *I*, (6), 689.

Blaikie, N. (2018). Confounding issues related to determining sample size in qualitative research. *I*, (5), 635.

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Q*, 201.

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Q*, (9), 1263.

Bragaru, M., van Wilgen, C. P., Geertzen, J. H., Ruijs, S. G., Dijkstra, P. U., & Dekker, R. (2013). Barriers and facilitators of participation in sports: A qualitative study on Dutch individuals with lower limb amputation. *P*, (3), e59881.

Coenen, M., Coenen, T., Stamm, A., Stucki, G., & Cieza, A. (2012). Individual interviews and focus groups with patients with rheumatoid arthritis: A comparison of two qualitative methods. *Q*, (2), 359.

Constantinou, C. S., Georgiou, M., & Perdikogianni, M. (2017). A comparative method for themes saturation (CoMeTS) in qualitative interviews. *Q*, (5), 571.

Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: Examples in the geography of health. *S*, 0(7-8), 1001.

Carlsen, B., & Glenton, C. (2011). What about N? A methodological study of sample-size reporting in focus group studies. *B*, (1), 1.

Corbin, J., & Strauss, A. (2014). *B*. Thousand Oaks, CA: Sage Publications, Inc.

Emmel, N. (2015). Themes, variables, and the limits to calculating sample size in qualitative research: A response to Fugard and Potts. *I*, (6), 685.

Francis, J., Johnson, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M., & Grimshaw, J. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *P*, (10), 1229.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *T*, (9), 1408.

Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. (6), 669.

Grady, M. P. (1998). *Q*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Goulding, C. (2005). Grounded theory, ethnography and phenomenology: A comparative analysis of three qualitative strategies for marketing research. *E*, (3/4), 294.

Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *J*, (1), 2.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *F*, (1), 59.

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *P*, (5), e0232076.

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *F*, (1), 3.

Glaser, B. G., & Strauss, A. L. (1967). *T*. Chicago, IL: Aldine Transaction.

Hammersley, M. (2015). Sampling and thematic analysis: A response to Fugard and Potts. *I*, (6), 687.

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Q*, (4), 591.

Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in focus group research. *Q*, (10), 1483.

Hagaman, A. K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? *S*, (1), 23.

Jackson, D., Daly, J., Davidson, P., Elliott, D., Cameron-Traub, E., Wade, V., Chin, C., & Salamonson, Y. (2000). Women recovering from first-time myocardial infarction (MI): A feminist qualitative study. *J*, (6), 1403.

Jackson, M., Harrison, P., Swinburn, B., & Lawrence, M. (2015). Using a qualitative vignette to explore a complex public health issue. *Q*, (10), 1395.

Jassim, G. A., & Whitford, D. L. (2014). Understanding the experiences and quality of life issues of Bahraini women with breast cancer. *S*, 189.

Kerr, C., Nixon, A., & Wild, D. (2010). Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *E*, (3), 269.

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *A*, (1), 26-46.

Lowe, A., Norris, A. C., Farris, A. J., & Babbage, D. R. (2018). Quantifying thematic saturation in qualitative data analysis. *F*, (3), 191.

Morse, J. M. (2015). "Data were saturated . . ." . *Q*, (5), 587.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *J*, (1), 11.

Marshall, V., & Long, B. C. (2010). Coping processes as revealed in the stories of mothers of children with autism. (1), 105.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Q*, (13), 1753.

Namey, E., Guest, G., McKenna, K., & Chen, M. (2016). Evaluating bang for the buck: A cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. (3), 425.

O' Reilly, M., & Parker, N. (2013). "Unsatisfactory saturation" : A critical exploration of the notion of saturated sample sizes in qualitative research. *Q*, (2), 190.

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *I*, (5), 619.

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. *Q*, (4), 1893.

van Manen, M., Higgins, I., & van der Riet, P. (2016). A conversation with Max van Manen on phenomenology in its original sense. *N*, (1), 4.

Vandecasteele, T., Debyser, B., Van Hecke, A., De Backer, T., Beeckman, D., & Verhaeghe, S. (2015). Nurses' perceptions of transgressive behaviour in care relationships: A qualitative study. *J*, (12), 2786.

Viet-Thi, T., Porcher, R., Viet-Chi, T., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *J*, 71.

Urquhart, C. (2012). *G*. Thousand Oaks, CA: Sage Publications, Inc.

Note: Figure translations are in progress. See original paper for figures.