Sample Size Determination in Medical Education Research

Sara Shafian¹ ⬤, Payam Khazaeli², Maryam Okhovati³*

¹Department of Medical Education, Education Development Center, Kerman University of Medical Sciences, Kerman, Iran
²Pharmaceutics Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran
³Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

***Corresponding author:** Department of Medical Education, Education Development Center, Kerman University of Medical Sciences, Kerman, Iran. Email: mokhovati@kmu.ac.ir

Background

The quality of a study is based on the appropriateness of its method and tools and may be influenced by the suitability of the adopted sampling strategy. Researchers have to make sampling decisions at the beginning of the study planning, the most important of which is whether a sample would represent the total population. Experienced researchers begin with the total population and work until achieving an appropriate sample size. In contrast, less experienced researchers often work bottom-up; they determine the least number of respondents required to conduct research [1]. However, unless they pre-identify the total population, it is not possible for them to evaluate how the sample they selected represents the total population. The other important factor is the sample size. In other words, how many participants should be considered? This article discusses on, how the sample should be representative of the population? It also aims to provide strategies for sample size determination for researchers in medical education research in simple language.

Sample Size and Representation of the Intended Population

In addition to the necessity of a minimum number of cases to assess the relationships between subgroups, researchers must achieve a minimum sample size that exactly represents the intended population. According to size, does a large sample guarantee representation? In the first example, the researcher was able to interview a total sample of 450 female students without considering the male student population. Does a small size guarantee representation? The second falls into the trap of saying that 50% of those who expressed their ideas said that "*medical students be active in outpatient clinics for two hours every weekday,*" while 50% were only uni-gender students. A too-large sample may be inefficient, and a too-small sample may be unrepresentative (e.g., in the first example, the researcher may be going to interview 450 students, but this is impractical in action, or they may have interviewed only 10 students); this may be performed, but it may most likely not be the representative of the total population of 900 students.

Where simple random sampling is used, the sample size necessary to reflect the population value of a specific variable depends on both the population size and the level of heterogeneity within the population [1]. For populations with equal heterogeneity or variance, the larger the population, the larger the sample. For populations with equal size, the higher the heterogeneity in a specific variable, the larger the sample required. In the case of a heterogeneous population, a large sample is preferred. In the case of a homogeneous population, a smaller sample is possible. As long as a sample cannot represent the engaged population exactly, it creates a sampling error [2].

Sample Size Calculation

In addition to choosing the sampling method, calculating the sample size also plays a decisive role in research quality regarding publishing research findings or drawing conclusions. Suppose a young assistant professor in a medical sciences university has one month to evaluate and compare medical students' viewpoints, encompassing 900 people in total, on establishing an effective doctor-patient relationship in several

universities using a semi-structured interview. Given the time available, interviewing all the students will be impossible, so he must choose among the students. How will he do that?

If he is going to interview 200 students, is not this number too many? If he is going to interview 20 students, is not this number too few? If he is going to interview only men or women, can he provide an equitable image of this research? How can he express the percentage of students who lack the required skills? There are such decisions and problems in research that depend on several factors, including sampling.

In such situations, decisions will influence the sampling strategy used. According to the above example, consider the assumption that several students are required as a research sample. An issue that should be taken into account is the sample size. A question that amateur researchers often ask is how many people should be in their sample in a study. Although the question is pretty simple, there is no obvious or simple answer because the sample size depends on factors such as research objectives, questions, and design, the size, nature, and heterogeneity of the population from which the sample has been extracted, the required level of trust and confidence interval, the required level of accuracy (the least sampling error that must be tolerated), the statistical power, the representation of the intended population in the sample, the number of categories in the sample, the variability of the investigated factor, the number of variables under investigation in the research, the statistical tests that should be used, and the research nature (for example, quantitative, qualitative, and mixed-method) (2). We will discuss these factors in the following.

Sample Size Based on Research Approach

It is generally better to have a large sample in quantitative research because it provides higher reliability and uses more complicated statistics. Hence, if researchers use some statistical data analyses, Martin (2018) suggests a thirty-people sample size for each variable for quantitative research as the minimum. However, this number is very few, and we recommend significantly more numbers. The suggestion is that before any data collection, researchers should consider the types of relationships they intend to discover in their final sample subgroups. The number of variables the researchers determine to control in their analysis and the types of statistical tests they intend to carry out should specify their decisions about sample size before conducting the research. A predicted minimum number of 30 cases for each variable should be usually used as a "rule of thumb"; it means that an individual should be assured of having at least 30 cases for each variable. However, it is again emphasized that this estimate is very low (2).

In qualitative research, there is no formula to define the intended number for each larger or smaller unit of data collection (3). The sample size in qualitative research is typically small and non-random, aiming at achieving a detailed description of the intended phenomenon. For instance, a phenomenological study may use a sample of 1 to 10 participants, or a grounded theory study may use 10 to 60 participants (4). Qualitative researchers collect data to describe the phenomena under investigation based on the participants' perspectives. The sample size in qualitative research is influenced by various factors, including the study scope, subject nature, data quality, study design, etc. (5). The principal factor in sample size estimation in qualitative research is the principle of data saturation (6).

Sample Size Based on Required Statistical Power

The sample size also depends on the type of data analysis. In some statistical tests, larger samples are generally required. For example, if the chi-square statistic is going to be calculated, the requirement of this test is the aligned data; for example, consider two subgroups of stakeholders in a university of medical sciences, including 80 fifth-year medical students and 20 professors, and their answers to a question on a five-point scale (Table 1):

In this example, the total sample size is 80 people, considered a sample with an apparently reasonable size. However, six out of 10 answer cells (60%) consist of fewer than five cases. The chi-square statistic requires five or more cases in 80% of the cells (i.e., eight out of 10 cells). In this example, only 40% of the cells consist of more than five cases; therefore, even with a relatively large sample, the statistical requirements for reliable data still need to be met by simple statistics such as chi-square. Obviously, as far as we can, we should predict possible data distributions and see whether they impede appropriate statistical analysis. If the distributions make it impossible to calculate reliable statistics, the sample size should be increased, or the data should be interpreted cautiously due to reliability-related problems.

The point here is that a relatively large sample size may be required for each variable. In fact, Gorard (2003) proposes that we can start from the minimum sample size needed for each cell, multiply it by the number of cells, and then double the total number. In the example mentioned above, the minimum sample size includes six cases in each cell; we multiply it by the table's number of cells and then double it: $120 = 2 \times 10 \times 6$.

Table 1. Example: The fifth-year medical students should be active in outpatient clinics for two hours every weekday

Variables	Completely Disagree	Disagree	No Idea	Agree	Completely Agree
The fifth-year medical students	25	20	3	8	4
Professors	6	4	2	4	4

However, to ensure, we suggest 10 in each cell, and we will have at least a sample of 200=2 × 10 × 10. However, there is even no guarantee that the distributions are correct (7).

Sample Size Based on Research Method

The sample size is also determined somehow according to the study design. For example, in a survey study, a large sample is typically required, particularly if inferential statistics are going to be calculated. In ethnographic or qualitative research, it is more likely that the sample size is small. Borg and Gall (1979) propose that a correlational study needs a sample size of fewer than 30 cases; in causal-comparative and experimental methods, the sample size should not be fewer than 15 cases, and survey research should not have less than 100 cases in each major subgroup and 20 to 50 cases in each minor subgroup. They suggest that sample size calculation should begin by estimating the smallest cases in the smallest subgroup of the sample and vice versa (8). Thus, for the aforementioned example (900 medical students), if 5% of the sample should be male students, and this subsample should include 30 cases (e.g., for correlational research), then the total sample would be 600 = 0.05 ÷ 30. If 15% of the medical student sample should be women and the subsample should be 45 cases, the total sample should be 300 = 0.15 ÷ 45.

Sample Size Proportional to the Investigated Population

The size of a probability sample (e.g., random sample) can be determined by two methods: By the researcher being cautious and confident that the sample represents the population's more extensive features with a minimum number of cases or by using a table that shows the proper size represents a random sample for a given number of the larger population through a mathematical formula. One of these examples has been provided by Krejcie and Morgan (1970), indicating that if a researcher extracts a sample from a population of 30 people or fewer, he is recommended to select the total population as a sample. Krejcie and Morgan show that the fewer the number of cases in the population, the higher the proportion of that population in the sample and vice versa. They remind that with increasing the population, the proportion of the required population in the sample reduces and indeed remains stable at about 384 cases (9).

Sample Size Using Confidence Interval

In determining the sample size for a probability sample, not only the population size but also the margins of error it is going to tolerate should be taken into consideration. These are expressed based on confidence level and confidence interval. The confidence level, normally expressed as a percentage (usually 95% or 99%), is an index of how confident we can be (e.g., 95% or 99% of the time) that the responses are placed at a given time. The variation range of a confidence interval is the degree of variation or range of variation (e.g., ±1%, ±2%, or ±3%) that a person is intended to be confident about. For instance, the confidence interval is ±3% in many surveys, meaning that if a poll indicates that a political party has 52% of the votes, it can reach 49% (52-3%) or 55% (52+3%). The confidence interval is influenced by the sample size, the population size, and the percentage of the sample that gives the "correct" answer. Here, the 95% confidence level indicates that we can be 95% confident that this result will be in the 46% to 55% range, i.e., ±3 per cent. The confidence level is calculated statistically based on the sample size, the confidence level, and the percentages of a level under the normal distribution curve; for example, a 95% confidence level covers 95% of the distribution curve. Here, the sample size is rapidly reduced by increasing the population size. In general (but not always), the larger the population, the smaller the probability sample proportion. Moreover, the higher the confidence level, the larger the sample, and the lower the confidence interval, the higher the sample. In a usual sampling strategy, a 95% confidence level and a 3% confidence interval are used (2).

Sample Size Based on Variable Type

Bartlett et al. (2001) suggest that sample size is different for categorical variables (e.g., gender, education level) than for continuous data (e.g., scores in a test). In categorical data, we usually need larger samples than in continuous data. Regarding categorical and continuous variables, Bartlett et al. (2001) suggest that for categorical data, a margin of error of 5% is considered, while a margin of error of 3% is taken into consideration for continuous data. Here, for both categorical and continuous data, the population proportion is reduced with increasing the sample, and for continuous data, no difference exists in sample size for populations of 2,000 or larger. If categorical and continuous data are used, the researcher should usually choose the largest one (i.e., the sample size required for categorical data) (10).

Sample Size Based on Statistical Analysis

Bartlett et al. (2001) indicate that sample size should be different based on the statistics used and recommend that if multiple regressions are going to be calculated, "the proportion of observations [cases] to independent variables should not be lower than five." However, some statisticians suggest a 1:10 ratio, especially for continuous data, as the sample size should not be smaller for continuous data than for categorical data. They also suggest that in multiple regression: (a) For continuous data, if the number of independent variables is in the 5:1 ratio, the sample size should not be less than 111, and the number of regressors (independent variables) should not be higher than 22; (b) for continuous data, if the number of independent variables is 1:10, the sample size should not be less than 111, and the number of regressors (independent variables) should not be more than 11; c) for categorical data, if the number of independent variables is 1:5, the sample size should not be less than 313, and the number of regressors (independent variables) should not be higher than 62; (d) for categorical data, if the number of independent variables is 1:10, the sample size should not be less than 313 and the number of regressors (independent variables) should not be higher than 31. Bartlett et al. (2001) suggest that for factor analysis, a sample size of fewer than 100 observations (cases) should be considered the general rule. However, the size can reduce by up to 30 cases, and the ratio of a sample size to the number of variables ranges from 1:5 to 1:30 (10, 11).

In addition, Borg and Gall (1979) provided a formula-based approach or sample size determination, i.e., looking at the significance levels of correlation coefficients and then reading the sample sizes is usually required to show the level. For example, if the correlation significance level is 0.01, our required sample size is 10, or if the required correlation coefficient is 0.20, the sample size is 100. Still, an inverse ratio can be observed-the larger the sample population is, the smaller the required correlation coefficient can be considered significant (8).

Conclusion

With quantitative and qualitative data, the substantial requirement is that the sample represents a population from which it has been extracted. In a thesis concerning the life history of one expert in medical education, the sample population is one (n = 1). In a qualitative study of 30 faculty members of the surgical ward to investigate their teaching experiences with a virtual reality method, a sample size of 5 or 6 may be enough. In general, in qualitative research, some factors,

including the scope of the research question, influence the sample size; the more general and extensive the research question is, the slower it is to achieve data saturation. Another factor is the collected data quality, which in the way toward attaining high-quality data, may lead the researcher later to data saturation. In addition, the study design can also have a decisive role in sample size in qualitative research. However, in quantitative research or study of a heterogeneous population, e.g., for investigating several variables among medical students of universities, a larger sample should be selected to respect that heterogeneity. In fact, factors such as appropriateness to the research aim, suitability to the research question(s), and compatibility to the research focus determine how many individuals, groups, populations, etc., are needed. Sampling decisions may determine the nature, reliability, validity, trustworthiness, usefulness, and generalizability of the collected data and, indeed, the way of data collection.

Supplementary Material(s): is available here [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Conflict of interests: There was no conflict of interest.

Ethical approval: Not applicable.

Funding/Support: Not applicable.

References

1. Bailey K. Methods of social research: Simon and Schuster. 4th ed. New York: Free Press; 2008.
2. Martin J. Research Methods in Education 8th edition edited by Louis Cohen, Lawrence Manion and Keith Morrison. Research methods in education. New York: Routledge; 2018: 323-33.
3. Yin RK. Qualitative research from start to finish. New York: Guilford Pub; 2015.
4. Starks H, Trinidad SB. Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory. Qual Health Res. 2007 Dec;17(10):1372-80. doi: 10.1177/1049732307307031. [PMID: 18000076]
5. Morse JM. Determining sample size. Thousand Oaks, CA; Sage Pub; 2000: 3-5.
6. Hewing H. Conducting research in conversation: A social science perspective. Abingdon, Oxon: Routledge; 2011.
7. Gorard S. Quantitative methods in social science research. London, UK: A&C Black; 2003.
8. Borg WR, Gall MD. Educational research: An introduction. British Journal of Educational Studies. 1984; 32(3): 274. doi: 10.2307/3121583.
9. Krejcie RV, Morgan DW. Determining sample size for research activities. Educ Psychol Meas. 1970;30(3):607-10. doi: 10.1177/001316447003000308.
10. Bartlett JE, Kotrlik JW, Chadwick C. Organizational research: Determining appropriate sample size in survey research. Information Technology, Learning, and Performance Journal. 2001;19(1):43-50.
11. Tabachnick BG, Fidell LS, Ullman JB. Using multivariate statistics. 5th ed. Boston, Massachusetts, US: Allyn & Bacon/Pearson Education; 2007.