Health Notions
DOI: http://dx.doi.org/10.33846/hn100401
EDITORIAL
ISSN 2580-4936
Volume 10 Number 4, April 2026

Interpreting Frequency Tables and Graphs in Health Research: Why the Largest Proportion Is Not Always the Main Finding

Heru Santoso Wahito Nugroho[1], Joel Rey Ugsang Acob[2]
[1]Department of Health, Poltekkes Kemenkes Surabaya, Surabaya, Indonesia; [2]Department of Nursing, Visayas State University, Baybay, Philippines
Correspondence: **Heru Santoso Wahito Nugroho:** Jl. Pucang Jajar Tengah 56 Surabaya, Indonesia; heruswn@gmail.com

Frequency tables and graphs are among the most frequently used methods for presenting categorical variables in health research [1-6]. They appear simple: researchers count how many participants belong to each category and then present the corresponding frequencies or percentages. However, although the statistical procedure is straightforward, the interpretation of these distributions requires careful consideration [7].

A common approach is to interpret the category with the largest proportion as the main finding [8-10]. This approach is generally appropriate for variables that are neutral descriptive characteristics, such as occupation, educational background, learning style, residence, or preferred source of health information [7]. For example, if the learning styles of health cadres are distributed as auditory = 21%, visual = 75%, and kinesthetic = 4%, it is reasonable to state that the majority of health cadres (75%) had a visual learning style. In this situation, the largest category is the most meaningful descriptive characteristic.

However, this principle should not automatically be applied to variables representing health status, disease conditions, or health risks. For these variables, the largest category may represent the desirable or normal condition and may therefore conceal an important public-health problem. In such circumstances, interpretation should focus on the proportion of participants experiencing an adverse health condition or possessing a relevant risk factor, even when that category is not the largest. This distinction is important because statistical description and public-health interpretation are not necessarily identical. A frequency table tells us how observations are distributed; interpretation should additionally tell us what that distribution means from a health perspective [7].

For a neutral variable, the main descriptive question is usually: Which category is most common? [7]. Consider the following hypothetical distribution of health cadres according to learning style (Table 1). The appropriate interpretation is: The majority of health cadres (75%) had a visual learning style. This interpretation is logical because visual learning style is not inherently a disease, health problem, or undesirable condition. The purpose of the analysis is primarily to describe the characteristics of the population. Therefore, the category with the largest proportion naturally becomes the focus of interpretation. The same principle can be applied to variables such as occupation, education, marital status, type of health insurance, or preferred health-information source when these variables are being treated simply as descriptive characteristics. For example, if 68% of patients are farmers, the researcher may state: Most patients (68%) were farmers. There is no statistical or epidemiological reason to interpret the 12% office workers or 8% unemployed participants as the principal finding merely because those categories may have particular contextual relevance.

Table 1. Distribution of learning style

Learning style	Frequency	Percentage
Auditory	42	21
Visual	150	75
Kinesthetic	8	4

Table 2. Distribution of nutritional status

Nutritional status	Frequency	Percentage
Overweight	4	2
Normal	154	77
Underweight	42	21

Table 3. Distribution of nutritional status

Maternal age at pregnancy	Frequency	Percentage
<20 years	36	18
20-35 years	144	72
>35 years	20	10

The situation changes when the variable represents a health condition. Suppose the nutritional status of 200 adolescents in Village A is distributed in Table 2. If the researcher simply follows the "largest percentage" rule, the interpretation would be: The majority of adolescents (77%) had normal nutritional status. Although this statement is statistically correct, it may be epidemiologically incomplete [7]. A more health-oriented interpretation would be: Underweight remained a substantial nutritional problem among adolescents in Village A, affecting 21% of the population. The reason is that 21% is not merely a small descriptive category. It represents one in every five adolescents experiencing an adverse nutritional condition. Even though normal nutritional status constitutes the largest category, the 21% underweight prevalence may represent an important public-health concern requiring attention. Thus, the researcher should not allow the 77% normal category to overshadow the health problem represented by the 21% underweight category. Importantly, describing a prevalence as "high" or "serious" should ideally be supported by an appropriate epidemiological or programmatic reference value. Nevertheless, a prevalence of 21% clearly warrants attention because of the substantial number of adolescents affected.

The key question is not always "which category is the largest?" This distinction can be summarized by changing the question used during interpretation. For a neutral variable, ask: Which category is the most common? For a health-status variable, ask: What proportion of the population has an adverse health condition? For a risk-factor variable, ask: What proportion of the population is exposed to the risk factor? These questions produce different interpretations from the same type of frequency table. The statistical procedure remains identical. What changes is the meaning assigned to the categories. This is particularly important in public-health research because the largest category is frequently the *normal* or *healthy* category. If researchers always focus on the largest percentage, they may inadvertently emphasize the healthy majority while underemphasizing a smaller but epidemiologically important subgroup.

The same logic applies to variables representing risk factors. Consider maternal age at pregnancy as an example (Table 3). Suppose 200 pregnant women are categorized according to whether their age at pregnancy is considered too young, within the recommended range, or too old according to the operational definition used in the study. A purely descriptive interpretation would be: Most pregnant women (72%) were within the recommended maternal age range. Again, this statement is correct, but it does not necessarily communicate the most important health-risk finding. For a public-health interpretation, attention should instead be directed toward the exposed groups: A considerable proportion of pregnant women experienced age-related pregnancy risk, with 18% becoming pregnant at an age considered too young and 10% at an age considered too old. The two risk categories can also be combined when scientifically justified: Overall, 28% of pregnant women were in age groups considered potentially at risk because of maternal age at pregnancy. This interpretation immediately communicates the public-health relevance of the distribution: more than one in four pregnancies occurred outside the predefined age range considered optimal in the study. The important point is that the 72% "recommended age" category should not automatically become the central finding simply because it is the largest category. The 18% and 10% categories represent exposures that may have implications for maternal and child health.

A frequency table has at least two levels of interpretation. The first level is statistical description: "Twenty-one percent of adolescents were underweight." The second level is substantive or public-health interpretation: "Underweight affected 21% of adolescents, indicating a substantial nutritional problem that warrants attention." The first statement describes the data. The second explains why the number matters. This distinction is particularly important in manuscripts because readers should not be required to determine the public-health implications themselves. Researchers should interpret the findings while remaining appropriately cautious.

However, interpretation should not exaggerate the evidence. A researcher should avoid automatically labeling a prevalence as "very high," "epidemic," or "serious" unless an accepted threshold, guideline, comparison population, or epidemiological criterion supports that conclusion. For example, instead of writing: "Only 21% were underweight." A more appropriate interpretation may be: "Underweight affected 21% of adolescents, representing a substantial proportion of the population and an important nutritional concern." This wording avoids minimizing the problem while also avoiding an unsupported claim about severity.

The same principle applies to risk factors. A researcher should first establish the operational definition of the risk category. For maternal age, for example, the categories "too young" and "too old" should not be introduced arbitrarily. The study should specify the criterion used to define these categories, based on an appropriate guideline or established epidemiological definition. The interpretation can then be expressed as: "Using the study's predefined age criteria, 18% of pregnant women were classified as being in the younger high-risk age group and 10% in the older high-risk age group." This is more scientifically defensible than simply stating: "Twenty-eight percent of mothers were too young or too old." The former makes clear that the classification depends on a predefined criterion.

The fundamental lesson is that frequency analysis is descriptive, but interpretation is contextual. The same percentage can have very different meanings depending on what the category represents. A category representing a preference, occupation, or learning style generally describes who or what is most common. A category representing disease, malnutrition, disability, inadequate health behavior, or exposure to a risk factor describes how much of a health problem or risk exists in the population. Consequently, researchers should move beyond the mechanical rule: "Interpret the largest percentage," and adopt a more appropriate rule: "Interpret the category that is most meaningful for the research question and, for health-related variables, explicitly communicate the magnitude of adverse health conditions and risk exposures." This approach makes statistical reporting more closely aligned with epidemiological reasoning and public-health decision making.

Accurate interpretation of frequency tables and graphs is not merely a matter of wording or presentation; it is a fundamental component of scientific research and publication [11-18]. In health research, an incorrectly interpreted distribution can transform accurate statistical information into a misleading message, particularly when a researcher focuses only on the largest proportion while overlooking a smaller but important proportion representing a disease, health problem, or risk exposure. Misleading interpretation may subsequently produce misleading conclusions, and misleading conclusions can lead to inappropriate scientific judgments, misguided public-health decisions, and ineffective allocation of resources. In the context of health care and public health, such errors may ultimately have consequences for interventions, policies, and the populations they are intended to protect. Therefore, researchers have a scientific and ethical responsibility not only to calculate frequencies and percentages correctly, but also to interpret them according to the substantive meaning of each variable and its implications for health. An error in interpretation can lead to misleading information [19]; misleading information can lead to misleading conclusions [20]; misleading conclusions can lead to wrong decisions [21] and wrong decisions in health can cause harm [22].

In conclusion, interpreting frequency tables and graphs in health research requires more than identifying the category with the largest percentage. For neutral descriptive variables such as occupation, learning style, or preferred information source, the largest proportion generally provides the most appropriate summary of the population's dominant characteristic. For variables representing health status, disease, or health risk, however, the researcher should deliberately identify and communicate the proportion experiencing an adverse condition or exposed to a relevant risk factor. The essential principle is therefore simple: for neutral variables, describe what is most common; for health-status and risk variables, emphasize what represents a health problem or meaningful exposure. Good interpretation transforms a frequency distribution from a collection of percentages into a meaningful statement about the health of a population.

Ethical consideration, competing interest and source of funding

-Although this editorial did not require formal ethical approval, all ethical principles were consistently upheld.
-There is no conflict of interest related to this publication.
-Source of funding is authors.

REFERENCES

1. Nugroho HSW, Santosa BJ. Misleading use of the terms of univariate and bivariate analysis in health research. Health Notions. 2019 Aug 31;3(8):352-6.
2. Kannan S, Dongare PA, Garg R, Harsoor SS. Describing and displaying numerical and categorical data. Airway. 2019 May 1;2(2):64-70.
3. Nugroho HSW. Cangkir Santi 50: Mengenal data kategorik [Internet]. Facebook. 2025 Nov 11 [cited 2026 Jan 3]. Available from: https://www.facebook.com/reel/1491544275449826
4. Suparji, Nugroho HSW, Martiningsih W. Tips for distinguishing nominal and ordinal scale data. Aloha International Journal of Multidisciplinary Advancement (AIJMU). 2019;1(6):133-5.
5. Nugroho HSW. Analisis data secara deskriptif untuk data kategorik. Ponorogo: Forum Ilmiah Kesehatan (FORIKES); 2014.
6. Mishra P, Pandey CM, Singh U, Gupta A. Scales of measurement and presentation of statistical data. Annals of Cardiac Anaesthesia. 2018 Oct;21(4):419.
7. Nugroho HSW. Cangkir Santi 3: Hati-hati menafsirkan data persentase [Internet]. Facebook. 2025 Aug 19 [cited 2026 Jan 3]. Available from: https://www.facebook.com/reel/1487784492242902
8. Suharto A, Nugroho HSW, Santosa BJ. Metode penelitian dan statistika dasar. Bandung: Media Sains Indonesia; 2022.

9. Spence I, Lewandowsky S. Displaying proportions and percentages. Applied Cognitive Psychology. 1991 Jan;5(1):61-77.
10. Pallauta JD, Gea MM, Arteaga P, Valenzuela-Ruiz SM. Secondary school students' interpretation of the frequency table. InBridging the Gap: Empowering y Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022). https://doi. org/10.52041/iase. icots11. t2c2 2022.
11. McGrath TA, McInnes MD, van Es N, Leeflang MM, Korevaar DA, Bossuyt PM. Overinterpretation of research findings: evidence of "spin" in systematic reviews of diagnostic accuracy studies. Clinical Chemistry. 2017 Aug 1;63(8):1353-62.
12. Kotronoulas G, Miguel S, Dowling M, Fernández-Ortega P, Colomer-Lahiguera S, Bağçivan G, Pape E, Drury A, Semple C, Dieperink KB, Papadopoulou C. An overview of the fundamentals of data management, analysis, and interpretation in quantitative research. InSeminars in oncology nursing 2023 Apr 1 (Vol. 39, No. 2, p. 151398). WB Saunders.
13. Marks M, Rodrigues JN. Correct reporting and interpretation of clinical data. Journal of Hand Surgery (European Volume). 2017 Nov;42(9):977-9.
14. Francis G. Equivalent statistics and data interpretation. Behavior research methods. 2017 Aug;49(4):1524-38.
15. Islam M, Hossain A, Hasan R, Soton AR. The methods, benefits and problems of the interpretation of data. In2020 International Conference on Information Science and Communications Technologies (ICISCT) 2020 Nov 4 (pp. 1-5). IEEE.
16. Abdi G, Jain M, Barwant M, Tendulkar R, Tendulkar M, Tariq M, Amir A. Unveiling the dynamic role of bioinformatics in automation for efficient and accurate data processing and interpretation. InAdvances in Bioinformatics 2024 Feb 6 (pp. 279-319). Singapore: Springer Nature Singapore.
17. Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. Annals of Cardiac Anaesthesia. 2019 Jul;22(3):297.
18. Queiroz T, Monteiro C, Carvalho L, François K. Interpretation of statistical data: The importance of affective expressions. Statistics Education Research Journal. 2017;16(1):163-80.
19. Alm-Kruse K, Kramer-Johansen J. Addressing data misinterpretation in recent meta-analysis. Journal of Intensive Care Medicine. 2025 Jun;40(6):691-2.
20. Tsuda T, Miyano Y, Yamamoto E. Mistaken information can lead only to misguided conclusions and policies: a commentary regarding Schüz et al.'s response. Environmental Health. 2023 Sep 2;22(1):62.
21. Jobe AH. Large data and the risks of misleading conclusions. Journal of Pediatrics. 2017 Jul;186:7-9.
22. Sameera V, Bindra A, Rath GP. Human errors and their prevention in healthcare. Journal of anaesthesiology, clinical pharmacology. 2021 Oct 12;37(3):328.