Neumol Pediatr 2019; 14 (3): 122 - 125

HOW TO READ AND GENERATE SCIENTIFIC PUBLICATIONS. ROLE AND DEFINITION OF VARIABLES IN AN INVESTIGATION: THE PROMINENCE THEY DESERVE

Kins. MPh. Karla Yohannessen V.[1,2], Eng, MSc. Mauricio Fuentes A.[3]

1 Assistant Professor, Department of Pediatrics and Child Surgery, School of Medicine, Universidad de Chile

2 Assistant Professor, Environmental Health Program, School of Public Health, Universidad de Chile

3 Assistant Professor, Biostatistics Program, School of Public Health, Universidad de Chile

ABSTRACT

When formulating a quantitative research question, implicitly the terms or variables included in it are being considered. The definition of the variables is necessary for the people who review or read the research to understand what the researcher is referring to when he/she mentions them, to make sure that they can be observed or measured, to be able to compare with similar investigations, and to analyze and adequately communicate the results. This article aims to highlight the importance of defining the variables in the process of an investigation.

Keywords: variable, measurement, definition, operationalize.

INTRODUCTION

In the article The Question, the Initial Step in Research or in the Search for Information (1), the measurement of variables is mentioned briefly as part of the scientific method process. However, in this manuscript we will highlight the importance of identifying and operationalizing variables in an investigation, whether in the field of clinical or epidemiological studies, from the research project to the dissemination of its results.

Starting from the most essential: What is a variable?

A variable is a characteristic or property of the subject under study and can be measured or quantified in one or several ways (2,3). An essential characteristic is that it must be variable, even if it sounds redundant, which means that it can vary or take different values in the unit of analysis to be studied (4). In research projects or in scientific communications, the variables of interest are always present, which are those over which learning is sought (3).

So, data from any investigation are the values of the variables of interest obtained or measured in the unit of analysis, which could correspond to people, patients, laboratory samples, units within a hospital, hospitals, districts, among many others.

Where do we identify variables in an investigation?

Variables are present in the problem to investigate from the very beginning and could correspond to the keywords that start a bibliographic search on the subject. Once the researcher has delved into the problem to be investigated, he is able to define a research question, which will be formulated in relation to the main variables of interest. Then, the formulation of the main and specific objectives will be based on the research question and the variables present in it, and in the case of specific objectives, these could include other sets of variables that help describe the context or that will contribute to understand the problem to study. Sometimes, research questions aim to study associations, relationships or comparisons between variables, so a research hypothesis would have to be formulated that also includes the main variables.

Correspondence:
Karla Yohannessen V.
Department of Pediatrics and Child Surgery, School of Medicine, Universidad de Chile. Independencia 1027, Independencia, Santiago, Chile.
+562 29786539
karlayohannessen@med.uchile.cl

The stages of planning (methodology), development, analysis and communication of the results and conclusions of the research will include the variables or sets of variables contained in the question, objectives and research hypotheses.

How and where are variables defined in an investigation?

Variables generally have a theoretical definition that corresponds to the meaning of the concepts being studied. Usually this definition is reported in the theoretical framework or rationale of a research project or in the introduction if it is a scientific communication. However, if the scientific communication is published in a journal of the specific area of the subject under investigation, there may be no need to include a theoretical definition.

On the other hand, the operational definition of the variables corresponds to the way in which these will be measured, specifying the procedures necessary for the identification of a concept in observable and measurable terms (2,4), how they will be recorded and what they represent, indicating its dimensions. Usually this definition is present in the methods section, either in the research project or the scientific communication.

As the operational definition of the variables implies having refined the question, objectives and theoretical framework, the ideal is to use definitions that have already been used and validated by other researchers; this will allow, on the one hand, to keep in mind the difficulties and limitations in the measurement reported and, on the other hand, to compare the results with other investigations.

What aspects should be considered when operationalizing variables?

One of the first aspects to consider when operationalizing variables from a study is the difficulty implied in its definition. Some variables are simple to define and do not represent major difficulties in conceptualization or measurement. These have been called objective variables (2), due to their clear definition, simple interpretation and the availability of a known and good instrument to measure them. Examples of these variables are weight, height, red blood cell count, among others. However, there are subjective variables, which are more difficult to define because they do not have a definition or a universally accepted form of measurement (2), such as pain intensity, quality of life, mental health symptoms, among others. In both cases, the operational definition of the variable should include the form and instrument that will be used to perform the measurement and, especially for subjective variables, a description of the reliability and validity of the measuring instrument, if available.

The operational definitions should be clear, unambiguous and should provide for all possible situations. For example, if it is of interest to know the number of asthmatic children's visits to the Emergency Room, the possible answers "none", "0", "I do not remember" or "I do not know", should not be left out, since if these alternatives are left out, it could be assumed that no child came in with an urgent consultation, since the person who answered the question did not know that information, generating an error in the measurement of the variable. For this reason, it is important to operationalize variables in the planning stage of the study, since once the study has begun it will be more difficult to correct errors in the definitions.

For the operationalization of some variables, approximate definitions are used (2). For example, if we wanted to know the exposure to tobacco smoke in children with chronic respiratory disease, a simple and economical way would be to use a questionnaire applied to the mother or responsible adult, providing an approximation of the child's exposure, that may very well not be as precise as the researcher would like, however, would bring him closer to the unknown reality and provide information so he can compare his results with other studies that used the same questionnaire. In this same example, if the researcher has more resources (economic, personal, time), he could use a urine cotinine measurement, which would deliver more precise values of exposure to tobacco smoke, although it could be difficult to get the urine sample or not be accepted by the responsible adult.

The previous example gives two fundamental aspects prior to the operational definition of the variables. The first is that all possible ways of measuring or quantifying the variable of interest should be considered, and the second is that the selection of the measurement form will be related to the economic resources, personnel and time available, feasibility and acceptability of the measurement form. Finally, the operational definition of the variable will include the description of the chosen measurement form.

In many cases, the operational definition of the variable is complex, because there is no single variable that, by itself, expresses the complexity of the phenomenon of interest. In this situation, it is advisable to use different variables where each one evaluates a different aspect of the phenomenon (including objective and subjective aspects), and from these generate the variable of interest. This will require that the operationalization of the complex variable includes the definition of those variables that compose it and an explanation of how it will be built. Classic examples of variables or complex phenomena of interest are socioeconomic status, quality of life, pain intensity, among others.

Another aspect to consider in the operational definition of the variables of interest, is the description of how the measurement will be obtained, that is, detail if there will be personnel to carry out the measurements for the study, if it will receive some type of training and if there will be any measurement protocol. For example, the value of a blood pressure measurement may vary depending on the position of the subject, previous activity and even the time of day, therefore, in this case a protocol should be made that includes whether to consider any specific position, a resting time and the preference of some time of day to perform the measurement. On the other hand, a database with variables and measurements already made or of hospital records could also be used, in this case, this should also be described in the operational definition and commented as a possible limitation of the study, since the variable could not be defined previously by the researcher and also did not have a planning in its measurement and recording form.

Finally, the variables that must be defined operationally and in greater detail in the methods section are those of greatest interest for the research. If it is an investigation that aims to describe clinical characteristics of subjects with a specific disease, those "characteristics" that are the variables of interest

should be defined in detail. On the other hand, if the research intends to estimate associations (statistics or causes) between variables, the operational definition in detail of each variable that is intended to be associated should be included. Obviously, the context variables and those that contribute to the understanding of the problem under study must also be defined operationally, with a level of detail that will depend on their importance for the investigation.

How should I organize variables for database registration?

Before starting with the data collection or measurements, it is very desirable to build a database and identify what type of variables will be recorded. A classic and very useful classification is, according to its scale of measurement, quantitative and qualitative. Table 1 details this classification, a sub-classification and examples for each (2,5,6). The measurement scale of a variable will determine the statistical analysis that can be performed (4).

Most of the time, the operational definition of the variable has the measurement scale implicit, however, sometimes there may be several possibilities (2). For example, current tobacco consumption could be a nominal variable (yes-no), ordinal (non-smoker-occasional smoker-moderate smoker-heavy smoker), discrete (No. of daily cigarettes) or continuous (grams of nicotine daily). Note that the categories or values for the current tobacco consumption variable would not be the same if the variable were life-long tobacco consumption or exposure to tobacco smoke. In general, it is recommended to plan the data collection and registration form of the variable on a continuous scale, since it provides more information on the variability and from this register it could transform the variable into discrete or ordinal categories, while the reverse process is not possible to perform (2).

An important aspect to consider in the operational definition of the variable and especially in its registration form, is the unit of measurement. It should be borne in mind that several of the sub-classifications and examples in Table 1 by definition do not have unity, but if the variable has unity it must be made explicit, mainly to have a homogeneous record, a correct analysis and an adequate interpretation. For example, it is not the same to record the size in centimeters or meters, and the weight in kilograms or grams.

Why is identifying and operationalizing variables useful?

Identifying and defining variables in an operational way is an unavoidable task in the full development of an investigation, from the initial question to the report of the results or the conclusions of the scientific communication. Its usefulness, according to the different phases of an investigation, is detailed in Table 2. As it can be observed, the greatest utility is in the research project stage which is the stage in which the study is planned to respond to the question and objectives of an investigation.

CONCLUSION

Research is a set of systematic, critical and empirical processes that apply to the study of a phenomenon. The variables are present throughout the process and development of an investigation; therefore, it is important to recognize that the identification and operational definition of the variables conform a valuable and transversal task in this process. In this way, various aspects that could ultimately influence the interpretation and communication of the results can be considered.

The authors declare no conflicts of interest.

Table 1. Classification of variables according to their measurement scale.

Classification	Sub-classification	Description	Examples
Qualitative	Nominal	Well defined non-numerical categories, which do not have a logical or natural order. Within this category we find the dichotomous variables (D), which can take 2 values.	Types of disease Types of symptoms Types of treatment Sex (D) Presence-Absence (D) Asthma Diagnosis (D)
	Ordinal	Well-defined non-numerical categories, which can be arranged in a logical or natural way.	Pain intensity (low-moderate-high) Asthma severity (mild-moderate-severe) Classification of nutritional status (low weight-eutrophic-overweight-obese)
Quantitative	Discrete	Take whole numbers as values	N. of children N. of hospital admissions N. of asthmatic crises
	Continuous	Take as values any numerical value	Weight (kg), Size (m), BMI (kg / m2) Temperature (° C) Forced vital capacity (ml)

Table 2. Usefulness of the identification and operationalization of the variables under investigation.

Research Project Phase
- Evaluate the feasibility of the study (identify problems in the definition and measurement of the variables) - Project economic and personnel resources for data collection (budget generation) - Identify sources of variability (consider measurement variations within and between subjects or observers) - Visualize possible measurement biases (take into consideration to avoid or prevent this bias) - Identify and include the measurement of confounding variables - Admit that there may be variables that cannot be measured in the study - Recognize the possible ethical conflicts of the type of measurement to be performed - Plan data collection - Plan the statistical analysis to respond to it or the research objectives
Data collection and measurement phase
- Identify possible errors in the data record - Protocolize the recording of measurements (eg define whether rounding will be used)
Analysis, results and scientific communication phase
- Analyze the data in a manner consistent with the research objective (s) - Understand the results and facilitate their interpretation - Compare the results obtained - Generate conclusions that respond to the research objective (s)

REFERENCES

1. Palomino M. MA. How to read and generate scientific publications. The question, the initial step in research or in the search for information. (Cómo leer y generar publicaciones científicas. La pregunta, paso inicial de la investigación o búsqueda de información.)Pediatr. Pulm. 2019; 14 (1): 9–11.
2. Argimón Pallás J, Jiménez Villa J. Clinical and epidemiological research methods.(Métodos de investigación clínica y epidemiológica) Elsevier España, S.L. 4th Edition, Spain, 2013.
3. Bennett JO, Briggs WL, Triola MF. Statistical Reasoning(Razonamiento estadístico) Pearson Education. 1st Edition, Mexico, 2011.
4. Hernández Sampieri R. Research Methodology(Metodología de la investigación) McGraw-Hill / Interamerican Editors. 5th Edition, Mexico, 2014.
5. Sheskin DJ. Levels of measurement. In: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall CRC. 3rd Edition, USA, 2004, page 31-34.
6. Gorgas García J, Cardiel López N, Zamorano Calvo J. Random variables. In: Basic Statistics for Science students. (Variables aleatorias. En: Estadística Básica para estudiantes de Ciencias) Department of Astrophysics and Atmospheric Sciences, Faculty of Physical Sciences, Comp University of Madrid. 1st Edition, Spain, 2011, page 63-78.