

# A Student's Guide to the Classification and Operationalization of Variables in the Conceptualization and Design of a Clinical Study: Part 1

Chittaranjan Andrade<sup>1</sup>

## ABSTRACT

Students without prior research experience may not know how to conceptualize and design a study. This article explains how an understanding of the classification and operationalization of variables is the key to the process. Variables describe aspects of the sample that is under study; they are so called because they vary in value from subject to subject in the sample. Variables may be independent or dependent. Independent variables influence the value of other variables; dependent variables are influenced in value by other variables. A hypothesis states an expected relationship between variables. A significant relationship between an independent and dependent variable does not prove cause and effect; the relationship may partly or wholly be explained by one or more confounding variables. Variables need to be operationalized; that is, defined in a way that permits their accurate measurement. These and other concepts are explained with the help of clinically relevant examples.

**Keywords:** Independent variable, dependent variable, confounding variable, operationalization of variables, hypothesis

**Key Message:** This article explains the following concepts: Independent variables, dependent variables, confounding variables, operationalization of variables, and construction of hypotheses.

In any body of research, the subject of study requires to be described and understood. For example, if we wish to study predictors of response to antidepressant drugs (ADs) in patients with major depressive disorder (MDD), we might select patient age, sex, age at

onset of MDD, number of previous episodes of depression, duration of current depressive episode, presence of psychotic symptoms, past history of response to ADs, and other patient and illness characteristics as potential predictors. These characteristics or descriptors are called

variables. Whether or not the patient responds to AD treatment is also a variable. A solid understanding of variables is the cornerstone in the conceptualization and preparation of a research protocol, and in the framing of study hypotheses. This subject is presented in two parts. This article, Part 1, explains what independent and dependent variables are, how an understanding of these is important in framing hypotheses, and what operationalization of a variable entails.

## Variables

Variables are defined as characteristics of the sample that are examined, measured, described, and interpreted. Variables are so called because they vary in value from subject to subject in the study. As an example, if we wish to examine the relationship between age and height in a sample of children, age and height are the variables of interest; their values vary from child to child. In the earlier example, patients vary in age, sex, duration of current depressive episode, and

<sup>1</sup>Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India.

**HOW TO CITE THIS ARTICLE:** Andrade C. A Student's Guide to the Classification and Operationalization of Variables in the Conceptualization and Design of a Clinical Study: Part 1. *Indian J Psychol Med.* 2021;43(2):177-179.

**Address for correspondence:** Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka 560029, India. E-mail: andradec@gmail.com

**Submitted:** 24 Jan. 2021  
**Accepted:** 24 Jan. 2021  
**Published Online:** 26 Feb. 2021



Copyright © The Author(s) 2021

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution- NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

**ACCESS THIS ARTICLE ONLINE**  
Website: [journals.sagepub.com/home/szj](http://journals.sagepub.com/home/szj)  
DOI: 10.1177/0253717621994334

response to ADs. Variables are classified as dependent and independent variables and are usually analyzed as categorical or continuous variables.

## Independent and Dependent Variables

*Independent variables* are defined as those the values of which influence other variables. For example, age, sex, current smoking, LDL cholesterol level, and blood pressure are independent variables because their values (e.g., greater age, positive for current smoking, and higher LDL cholesterol level) influence the risk of myocardial infarction. *Dependent variables* are defined as those the values of which are influenced by other variables. For example, the risk of myocardial infarction is a dependent variable the value of which is influenced by variables such as age, sex, current smoking, LDL cholesterol level, and blood pressure. The risk is higher in older persons, in men, in current smokers, and so on.

There may be a cause–effect relationship between independent and dependent variables. For example, consider a clinical trial with *treatment* (iron supplement vs placebo) as the independent variable and *hemoglobin level* as the dependent variable. In children with anemia, an iron supplement will raise the hemoglobin level to a greater extent than will placebo; this is a cause–effect relationship because iron is necessary for the synthesis of hemoglobin. However, consider the variables *teeth* and *weight*. An alien from outer space who has no knowledge of human physiology may study human children below the age of 5 years and find that, as the number of teeth increases, weight increases. Should the alien conclude that there is a cause–effect relationship here, and that growing teeth causes weight gain? No, because a third variable, age, is a confounding variable<sup>2-3</sup> that is responsible for both increase in the number of teeth and increase in weight. In general, therefore, it is more proper to state that independent variables are *associated with* variations in the values of the dependent variables rather than state that independent variables *cause* variations in the values of the dependent variables. For causality to be asserted, other criteria must be fulfilled; this is out of the scope of the present article,

and interested readers may refer to Schunemann et al.<sup>4</sup>

As a side note, here, whether a particular variable is independent or dependent will depend on the question that is being asked. For example, in a study of factors influencing patient satisfaction with outpatient department (OPD) services, patient satisfaction is the dependent variable. But, in a study of factors influencing OPD attendance at a hospital, OPD attendance is the dependent variable, and patient satisfaction is merely one of many possible independent variables that can influence OPD attendance.

## Importance of Variables in Stating the Research Objectives

Students must have a clear idea about what they want to study in order to conceptualize and frame a research protocol. The first matters that they need to address are “What are my research questions?” and “What are my hypotheses?” Both questions can be answered only after choosing the dependent variables and then the independent variables for study.

In the case of a student who is interested in studying predictors of AD outcomes in patients with MDD, treatment response is the dependent variable and patient and clinical characteristics are possible independent variables. So, the selection of dependent and independent variables helps defines the objectives of the study:

1. To determine whether sociodemographic variables, such as age and sex, predict the outcome of an episode of depression in MDD patients who are treated with an AD.
2. To determine whether clinical variables, such as age at onset of depression, number of previous depressive episodes, duration of current depressive episode, and the presence of soft neurological signs, predict the outcome of an episode of depression in MDD patients who are treated with an AD.

Note that in a formal research protocol, the student will need to state all the independent variables and not merely list examples. The student may also choose to include additional independent variables, such as baseline biochemical, psychophysiological, and neuroradiological measures.

## Importance of Variables in Framing Hypotheses

A hypothesis is a clear statement of what the researcher expects to find in the study. As an example, a researcher may hypothesize that longer duration of current depression is associated with poorer response to ADs. In this hypothesis, the duration of the current episode of depression is the independent variable and treatment response is the dependent variable. It should be obvious, now, that a *hypothesis can also be defined as the statement of an expected relationship between an independent and a dependent variable*. Or, expressed visually, (independent variable) (arrow) (dependent variable) = hypothesis.

It would be a waste of time and energy to do a study to examine only one question: whether duration of current depression predicts treatment response. So, it is usual for research protocols to include many independent variables and many dependent variables in the generation of many hypotheses, as shown in **Table 1**. Pairing each variable in the “independent variable” column with each variable in the “dependent variable” column would result in the generation of these hypotheses. **Table 2** shows how this is done for age. Sets of hypotheses can likewise be constructed for the remaining independent and dependent variables in **Table 1**. Importantly, the student must select one of these hypotheses as the primary hypothesis; the remaining hypotheses, no matter how many they are, would be secondary hypotheses. It is necessary to have only one hypothesis as the primary hypothesis in order to calculate the sample size necessary for an adequately powered study and to reduce the risk of false positive findings in the analysis.<sup>5</sup> In rare situations, two hypotheses may be considered equally important and may be stated as coprimary hypotheses.

## Operationalization of Variables

In **Table 1**, suicidality is listed as an independent variable and severity of depression, as a dependent variable. These variables need to be operationalized; that is, stated in a way that explains how they will be measured. **Table 3** presents three ways in which suicidality can be measured and four ways in which (reduction

TABLE 1.

### Independent Variables and Dependent Variables in a Study on Sociodemographic and Clinical Prediction of Response of Major Depressive Disorder to Antidepressant Drug Treatment

#### Independent Variables

- Age
- Sex
- Age at onset of major depressive disorder
- Number of past episodes of depression
- Past history of response to antidepressant drugs
- Duration of current depressive episode
- Baseline severity of depression
- Baseline suicidality
- Baseline melancholia
- Baseline psychotic symptoms
- Baseline soft neurological signs

#### Dependent Variables

- Severity of depression
- Global severity of illness
- Subjective well-being
- Quality of life
- Everyday functioning

in) the severity of depression can be measured. Now, each way of measurement in the “independent variable” column can be paired with a way of measurement in the “dependent variable” column, making a total of 12 possible hypotheses. In like manner, the many variables listed in **Table 1** can each be operationalized in several different ways, resulting in the generation of a very large number of hypotheses. As already stated, the student must select only one hypothesis as the primary hypothesis.

Much thought should be given to the operationalization of variables because variables that are carelessly operationalized will be poorly measured; the data collected will then be of poor quality, and the study will yield unreliable results. For example, socioeconomic status may be operationalized as lower, middle, or upper class, depending on the patient’s monthly income, on the total monthly income of the family, or using a validated socioeconomic status assessment scale that takes into consideration income, education, occupation, and place of residence. The student must choose the method that would best suit the needs of the study, and the method that has the greatest scientific acceptability.

TABLE 2.

### Combinations of Age with Dependent Variables in the Generation of Hypotheses

In patients with major depressive disorder who are treated with antidepressant drugs:

1. Older age is associated with less attenuation in the severity of depression.
2. Older age is associated with less attenuation in the global severity of illness.
3. Older age is associated with less improvement in subjective well-being.
4. Older age is associated with less improvement in quality of life.
5. Older age is associated with less improvement in everyday functioning.

TABLE 3.

### Possible Ways of Operationalization of Suicidality and Depression

Independent Variable: Suicidality	Dependent Variable: Severity of Depression
<ul style="list-style-type: none"> <li>• Item score on the HAM-D</li> <li>• Item score on the MADRS</li> <li>• Beck scale for Suicide ideation total score</li> </ul>	<ul style="list-style-type: none"> <li>• MADRS total score</li> <li>• HAM-D total score</li> <li>• HAM-D response rate</li> <li>• HAM-D remission rate</li> </ul>

HAM-D: Hamilton Depression Rating Scale, MADRS: Montgomery–Asberg Depression Rating Scale.

However, it is also permissible to operationalize the same variable in many different ways and to include all these different operationalizations in the study, as shown in **Table 3**. This is because conceptualizing variables in different ways can help understand the subject of the study in different ways.

Operationalization of variables requires a consideration of the reliability and validity of the method of operationalization; discussions on reliability and validity are out of the scope of this article. Operationalization of variables also requires specification of the scale of measurement: nominal, ordinal, interval, or ratio; this is also out of the scope of the present article. Finally, operationalization of variables can also specify details of the measurement procedure. As an example, in a study on the use of metformin to reduce olanzapine-associated weight gain, we may state that we will obtain the weight of the patient but fail to explain how we will do it. Better would be to state that the same weighing scale will be used. Still better would be to state that we will use a weighing instrument that works on the principle of moving weights on a levered arm, and that the same instrument will be used for all patients. And best would be to add that we will weigh patients, dressed in standard hospital gowns, after they have voided their bladder but before they have eaten breakfast. When the way in which a variable will be measured is defined, measurement of that variable becomes more objective and uniform

## Concluding Notes

The next article, Part 2, will address what categorical and continuous variables are, why continuous variables should not be converted into categorical variables and when this rule can be broken, and what confounding variables are.

## Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

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

## References

1. Rhodes AE, Lin E, and Streiner DL. Confronting the confounders: the meaning, detection, and treatment of confounders in research. *Can J Psychiatry* 1999; 44(2): 175–179.
2. Mamdani M, Sykora K, Li P, et al. Reader’s guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* 2005; 330(7497): 960–962.
3. Andrade C. Prenatal depression and infant health: the importance of inadequately measured, unmeasured, and unknown confounders. *Indian J Psychol Med* 2018; 40(4): 395–397.
4. Schunemann H, Hill S, Guyatt G, et al. The GRADE approach and Bradford Hill’s criteria for causation. *J Epidemiol Community Health* 2011; 65(5): 392–395.
5. Andrade C. The primary outcome measure and its importance in clinical trials. *J Clin Psychiatry* 2015; 76(10): e1320–e1323.