Learning Curve

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

Chittaranjan Andrade¹

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 sample. 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...
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 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.4

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. 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   | Dependent Variables          |
|-------------------------|------------------------------|
| Age                     | Severity of depression       |
| Sex                     | Global severity of illness   |
| Age at onset of major depressive disorder | Subjective well-being |
| Number of past episodes of depression | Quality of life |
| Past history of response to antidepressant drugs | Everyday functioning |
| Duration of current depressive episode | Baseline severity of depression |
| Baseline suicidality    | Baseline melancholia         |
| Baseline psychotic symptoms | Baseline soft neurological signs |

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 |
|-----------------------------------|--------------------------------------------|
| Item score on the HAM-D           | MADRS total score                           |
| Item score on the MADRS           | HAM-D total score                           |
| Beck scale for Suicide ideation total score | HAM-D response rate |
|                                   | HAM-D remission rate                        |

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 adequately measured, unmeasured, and unknown confounders. Indian J Psychol Med 2018; 40(4): 392–397.
4. Schulman 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.