

Title: Variables are not “worryables”

Choosing the variables to study

Chapter Summary:

In this chapter, we will look at variables that are needed to answer research questions. We will explore briefly the different types of variables, determining the variables to study, measuring variables and the importance of variables within the context of a research study.

At the end of this chapter:

You should be able to create a list of variables, indicate the type of each variable in the list, measurement units and measurement procedure for each variable, indicate who will measure, and justify choice of the variable.

You have defined your research question. You have a hypothesis in hand and are now ready to start research that will either prove or disprove your hypothesis. You have searched the literature, found evidence that indicates that it is worth the while to pursue your question. You have then gone about choosing an appropriate design for your study

Let us do a quick recap here.

- A research question is a SMART question.
- A hypothesis is a comment or a statement that we aim to prove or disprove through our research.
- Reliable answers to a research question are obtained through a study that is designed appropriately

How do you prove or disprove your hypothesis?

You have to collect information.

- You have to collect *pertinent information*
- You have to collect pertinent information *as appropriate measures*
- You have to collect pertinent information as appropriate measures and *by appropriate procedures*
- You have to collect pertinent information as appropriate measures, by appropriate procedures and *by appropriate persons/methods*
- You cannot claim The Right to Information Act here (maybe you can).

You may have to collect several pieces of information to reach from your question to addressing your hypothesis. Several pieces of the puzzle exist and have to be sorted out.

What is a variable?

The pieces of information that you collect are called variables.

Why are they called variables?

They are called variables because the information collected may vary between persons or within persons over a period of time.

Examples:

- Age may vary between persons; Age may vary for the same person if measured over time
- Blood Pressure may vary between persons; BP may vary for the same person if measured over time

Age is a variable- It can vary.

Blood Pressure is a variable- It can vary.

Variables can vary- there are different types of variables

Variables can be

- Dichotomous or nominal
- Discrete, Categorical or Ordinal
- Continuous, interval or ratio

Dichotomous Variables

These are variables with two levels- either/or. An example is a yes/no response. Gender – Male/Female is dichotomous.

Discrete Variables

These are variables that have more than two levels. For example, strongly disagree, disagree, agree, and strongly agree. The levels can be ranked in discrete variables; however, placing a value on the levels is often not possible.

Continuous variables

These are variables that exist in a continuum. Age is an example for a continuous variable. Height, weight are other examples of continuous variables.

In research, we often use the terms *dependent* variable and *independent* variable.

The dependent variable is a variable that may be influenced or is dependent on the independent variable.

Example:

Blood Pressure- dependent variable

Independent Variable- Age, Sex

The dependent variable blood pressure may be dependent on or influenced by the independent variables, age or sex. Thus, older people may have higher blood pressure. Blood Pressure may be higher in males than females.

The reverse does not hold true, a dependent variable cannot influence the independent variable. For example, a high BP cannot make you older or younger, or cannot change your gender.....

Choosing the variables to study

- Can use a trawler approach- thrown the net to collect as many variables as possible- some may be useful, some may not be useful to answer your question
- Can use a scientific focused approach-
 - Will collect variables that are pertinent to the question of interest.
 - Will not collect additional variables unless they add value to the question asked.
 - Will consider that each variable to be measured needs cooperation of the research subject and your time. If you will not use the information or the information is not needed, you are disrespecting the time of the study subject - bad ethics.

The trawler approach

- We are going to the subject so let us collect all information possible. Will help us avoid repeated visits for repeated information collection.
 - Problem: cannot be sure if information collected now will be relevant for a later time period
 - Problem: Study subject has to give more time, may put off the subject leading to refusals
 - Problem: Use of information for other studies does not often happen as different questions may need different levels of measurement
 - Problem: Need sustained training for data collection personnel
 - Problem: Need a larger data entry form
 - Problem: Need more data entry and a larger database
 - Problem: If will not use all that information after all that effort, why collect it-everyone is overworked as is!
- Example: We are going to those villages- let us collect all possible information from each household- I don't want to come back again for another study

How do you choose which variables to study?

- Personal experience
- Literature review
- Knowledge from Peers
- Knowledge from Experts
- Everyone collects it or ABC collected it, so I will also collect it- NOT GOOD. An example may be socioeconomic status, education, occupation, reported monthly income. Do not collect it unless it has a relationship with the question you are asking.

Measuring Variables

- Use standard, validated methods of measurement
- Identify standard methods through a literature review
- Example- How will you measure Hemoglobin? How will you measure diabetes? How will you measure blood pressure? What standard procedure and what standard units of measurement?

Who does the measurement?

- Anyone can measure anything- WRONG
- Some people can measure something- Clinical measurements by clinicians
- People can be trained to make measurements- consider training, duration, content, outcomes, certification as successfully trained
- Example: Who can read a biopsy as a variable?
- Document all training for study personnel- who got trained, who trained, what duration, content of training, outcomes of training, outcomes assessment, and certification.

Why are variables important?

The choice of wrong variables can introduce a lot of “noise” in the analysis and results. Wrongly characterized or measured variables can lead to loss of information.

For example:

- Collecting age as a discrete variable, 0-10, 11-20, 21-30, 31-40, 40 and above will not help us to present the mean, median or range of values.
- Collecting duration of illness in years or months for acute conjunctivitis- the condition often subsides in days so the unit of measurement has to be days.

You can save a lot of heartache if you plan your variables properly and well in advance of data collection

KEY QUESTIONS

1. What value does this variable add to my study?
2. Is this necessary?
3. Do I have proof of its utility?
4. Is it feasible to collect this information?
5. Who will collect and how will it be measured?
6. What will I lose if I don't study this variable?
7. Is this variable linked to other variables in the study?
8. Do they interact with other variables in any way? How does, if it does, this variable influence or impact other variables in the study?
9. MOST IMPORTANT-HOW WILL I USE THIS INFORMATION AND DOES THE METHOD OF ASCERTAINMENT ALLOW ME TO USE THIS INFORMATION APPROPRIATELY?