Designing a Study
Operational Definitions

The following chapter is excerpted from *Designing HIV/AIDS Intervention Studies: An Operations Research Handbook*, Andrew Fisher and James Foreit, 2002, Washington, DC: Population Council. (More on OR Handbook)

CHAPTER 6

OPERATIONAL DEFINITIONS

After formulating the study objectives and hypotheses and describing fully the study intervention, the next step in the research process is to define operationally the key variables and terms of the study. Operational definitions serve two essential purposes: (1) They establish the rules and procedures the research investigator will use to measure the key variables of the study, and (2) they provide unambiguous meaning to terms that otherwise might be interpreted in different ways. Every research proposal must include operational definitions of major variables and terms.

Operational Definitions of Variables

Suppose that a dependent variable of a study is knowledge about how HIV/AIDS is transmitted. Before this variable can be measured, it is necessary first to establish the operational procedures that specify how the measurement will be made and at the same time define what the researcher means by the words "knowledge about how HIV/AIDS is transmitted." This variable must be defined in terms of events that are **observable by the senses** and therefore measurable.

The observable events serve as an **indicator** of the variable, knowledge about HIV/AIDS transmission. Alone and by itself, knowledge is not observable by the senses. It is an abstract concept. You cannot touch knowledge, see it, smell it, taste it, or hear it. What is needed is an observable event that can be measured and that **indicates** knowledge. Usually, such an indicator of knowledge in an HIV/AIDS study is based on a series of questions. For example, you might ask a respondent, "Do you know how a person can become infected with AIDS?" "Please list all the ways you know a person can get AIDS." "Can a person get AIDS from a mosquito bite?" "Can HIV/AIDS be transmitted through a mother's breast milk?" Each of these questions indicates whether the respondent knows about certain aspects of HIV/AIDS transmission. Asking a question and hearing a response is an observable event that can be measured.

A research study might ask ten HIV/AIDS knowledge questions. Each time a respondent gives an answer that indicates knowledge about HIV/AIDS transmission, the researcher could record a score of one. Every time an answer is given that does not indicate knowledge about HIV/AIDS transmission, the researcher could record a score of 0. For each respondent, the researcher could then add the total number of correct answers to the ten questions and create a HIV/AIDS knowledge score. This score would range from 0 correct answers to ten correct answers. Persons with a score of 0 would be operationally defined as having no knowledge about HIV/AIDS transmission. Persons with a score of ten would be operationally defined as having a high level of knowledge about HIV/AIDS transmission. In your research proposal, the operational definition of knowledge might appear as:

Knowledge about HIV/AIDS transmission	=	The number of correct answers a respondent gives to ten questions on HIV/AIDS transmission.

This is not the only way the variable could be defined operationally. You might wish to establish categories of HIV/AIDS knowledge, distinguishing between those respondents who have high HIV/AIDS knowledge, medium knowledge, low knowledge, and no knowledge. Each of these levels is a category of the variable, and each category requires an operational rule that tells you how to assign any given respondent to the category. One way of operationally defining the categories might be as follows:

High knowledge	=	Correct responses to eight or more of the ten questions.
Medium knowledge	=	Correct responses to between four and seven of the ten questions.
Low knowledge	=	Correct responses to between one and three of the ten questions.
No knowledge	=	No correct answers to any of the ten questions.

Note that the four categories of the variable are **mutually exclusive**, that is, they do not overlap. According to the operational rules established, a person cannot be placed in the category "High Knowledge" and at the same time be placed in the "Medium," "Low," or "No" category. The categories are also **totally inclusive**. There are only four categories. There is no fifth, sixth, or seventh category that a respondent might fit into.

In some instances, you may not want to be quite so specific in defining the categories of a variable before data collection. Sometimes it is preferable to determine the category "cutting points" of a variable after data have been collected and the response distribution for the variable has been examined. As a general rule, it is best to have approximately an equal number of respondents in each category. Thus, in the example above, each of the four categories of the variable—knowledge about HIV/AIDS transmission—should have approximately 25 percent of the respondents in the study population.

If it is necessary to examine the response distribution of a variable before the procedures for establishing categories can be determined, then in the operational definition section of a study proposal the category names can be specified, but you should include a note indicating that each category will consist of approximately equal numbers of respondents.

All variables must have at least two or more categories, or they are not variables but instead are constants. Whenever you are operationally defining a variable, it is always better to divide the variable into many categories instead of just a few. In the examples given above, the variable knowledge about HIV/AIDS transmission ranges from 0 to ten. That range gives a total of 11 categories.

Subsequently, in the second example shown below, we collapsed these 11 categories into just four categories consisting of high, medium, low, and no knowledge. If we wanted to, we could go even further and collapse the four categories into just two:

Knowledge of HIV/AIDS transmission	=	A correct response to one or more of the ten questions.
No knowledge of HIV/AIDS transmission	=	No correct answers to any of the ten questions.

If you start with many categories, it is always easy to collapse these down to just a few. But do not make the mistake of starting with just a few categories, because subsequently you cannot expand them. Collapsing the categories of a variable is usually done after data collection has been completed and the frequency distribution of the variable has been examined. Sometimes it is possible to determine the categories of a variable on the basis of a good questionnaire pretest.

Examples of Operationally Defined Variables

Condom use	=	The reported use of a condom at the last act of intercourse.
Frequent condom use	=	The reported use of a condom during the last five or more acts of intercourse.
Peer educator performance	=	Any peer educator who holds at least one group meeting on HIV/AIDS per month or visits at least two homes of PLHA per month.
Modern village	=	Any village that has three or more of the following facilities: electricity, a government health clinic, a paved road within half a mile, a primary school, a bank, a post office, irrigation for 50 percent or more of the farmland.

Operational Definitions of Terms

Recall that a hypothesis is a statement about an expected relationship between two or more variables. Just as it is necessary to define variables operationally, it is also necessary to operationally define the terms that indicate the nature of the relationship between the variables. For example, in many hypothesis statements, you will find such terms as those shown below:

more than	greater than
less than	larger than
higher than	bigger than
lower than	smaller than

You are also likely to see in hypothesis statements such words as these:

safer	significant
acceptable	expanded
improved	increased

Each of these terms can have a variety of meanings, so each requires an operational definition for the research proposal. The basic problem with such terms as *more than* or *less than* or *increased* is that they suggest a comparison but do not indicate the standard for the comparison. We need to know how much more and how much less and increased by how much.

Suppose a study has the following simple hypothesis:

> A five-week, field-based training program will increase the knowledge about HIV/AIDS transmission among peer educators who have taken the program.

In this example, the training program is the independent variable. In the hypothesis, this variable is already defined, at least partially, as five weeks long and field-based. Knowledge about HIV/AIDS transmission is the dependent variable. We already have defined this variable as the number of correct responses to ten questions. What remains to be done is to define the term *increase*. If you do not define this term, you will find it impossible to know when the hypothesis has been proved or disproved. In other words, you need a standard of comparison that will tell you *increase by how much*. One way to define *increase* might be the following:

Increase = Among peer educators, a mean HIV/AIDS knowledge score on the post-training test that is significantly greater ($p < .05$) than the mean HIV/AIDS knowledge score of a control group of peer educators who did not participate in the training program.

Note that this operational definition not only tells us the meaning of increase but also gives us the procedures that will be used to measure the increase. The mean HIV/AIDS knowledge score of peer educators (in an experimental group) will be compared against the mean HIV/AIDS knowledge score of a control group. The hypothesis will be accepted only if the mean score of the peer educators in the experimental group is greater than and significantly different from the mean score of the control group. To be absolutely clear, we also should define the word *significantly*:

Significantly = A probability equal to or greater than .95 that the mean score of the peer educators in the experimental group is higher than the control group mean score.

To summarize, operational definitions establish the rules and procedures an investigator plans to use to measure and give meaning to variables and terms. The operational definition identifies indicators that are **observable events**. We must be able to ask a question, hear a response, see a behavior, record an action, and measure an attribute. The definition establishes categories for variables. The categories must be **mutually exclusive** and **totally inclusive.** Operational definitions also establish the **standard of comparison** the investigator will use to either accept or reject a hypothesis.

What To Do: Writing Operational Definitions

1. Write an operational definition for each variable on your list of independent and dependent variables.

2. Write an operational definition for each **term** (such as greater than, less than, increased, and significant) used to indicate the nature of the relationship between variables.

3. For each definition you write, ask yourself:

    - Are the rules and procedures for measuring the variables clear?
    - Have mutually exclusive and totally inclusive categories for the variables been established?
    - Is the standard of comparison clear for each term?