“Validity and Reliability in Research”

Chris Handley, MS, EMT-P, CPTC, Donor Services Coordinator, Transplant Resource Center of Maryland, Baltimore, MD, NATCO Research Committee Member

Whether you are planning a research project or interpreting the findings of someone else’s work, determining the impact of the results is dependent upon two concepts: validity and reliability. Essentially, validity entails the question, “does your measurement process, assessment, or project actually measure what you intend it to measure?”. The related topic of reliability addresses whether repeated measurements or assessments provide a consistent result given the same initial circumstances.

In research, validity has two essential parts: internal and external. Internal validity encompasses whether the results of the study (e.g. mean difference between treatment and control groups) are legitimate because of the way the groups were selected, data was recorded or analysis performed. For example, a study may have poor internal validity if testing was not performed the same way in treatment and control groups or if confounding variables were not accounted for in the study design or analysis. External validity, often called “generalizability”, involves whether the results given by the study are transferable to other groups (i.e. populations) of interest (Last, 2001). A study performed exclusively in a particular gender, racial, or geographic sub-group, such as white females in Appalachia, may not be applicable to Hispanic men in the northwest. It is through proper study design and strict protocol execution that high levels of validity, both internal and external, can be achieved. An important point to remember when discussing validity is that without internal validity, you cannot have external validity. Results of a poorly designed or executed study are not applicable to any population, in that particular study sample or otherwise.

A common threat to internal validity is reliability. Assuming the same initial conditions for a test assessment or process the test must provide the same result every time it is performed for it to be deemed reliable. For example, if a research protocol dictates that subjects must have their weight measured, the scale should provide the same weight if repeated measures were taken at the same time. Reliability is often at risk when assessments are taken over time, performed by different people or the assessments are highly subjective. In the weight example, if the subject had their weight taken monthly over a period of a year the scale may lose calibration over that time or differing clinic staff may read the scale differently over the study duration. Often physical assessment techniques, such as level of edema in a patient, are unreliable as they are highly subjective as they are “in the eye of the beholder”. As a researcher, you must ensure that these reliability errors are minimized so that if differences are seen in the data they can be attributed to the intervention and not to sloppy weight measurements.

It is with these concepts in mind that methods sections in protocols and journal articles provide such extensive detail related to how a study was designed and conducted. Threats to a study’s validity and reliability exist at almost every turn in the research process. No one researcher can see all the potential problems, so the team approach to the discussion of validity and reliability during the development of the study design, and creating and following study protocols can minimize the threats to validity and reliability.

Reference:

Last, J. (Ed.). (2001). *International epidemiological association. A dictionary of epidemiology* (4th ed.). New York: Oxford University Press.