

A Primer on Scores: What Counts?

To judge the effects of clinical interventions, researchers look for changes in certain key variables—better known as outcome measures. Some of the most familiar (and most important) outcome measures are dichotomous variables (so-called “0,1 variables”): They either happen or they don’t. Examples include heart attacks, strokes, and death. Other outcome measures can take on many values. Physiologic and laboratory measurements fall into this category (such as blood pressure, serum sodium levels, and CD4 counts), as do various functional status and symptom scales (such as the Glasgow Coma Scale to classify level of consciousness and visual analog scales to classify level of pain).

Over the past two decades, a new type of outcome measure has been increasingly used in clinical research: scores. A score is a composite measure—in other words, it is derived from several individual variables. A score may be the composite of multiple dichotomous variables, multiple physiologic and laboratory measurements, multiple scales, or any combination thereof. Scores are used primarily to measure multiattribute patient function (e.g., Mini-Mental Status Score is a metric for classifying the combined functions of orientation, computational ability, and short-term memory) or to predict risk for various outcomes (e.g., heart attack, breast cancer, or death).

Because they may summarize several different variables (which may have various weights), it can be difficult to know what a score really means. If the topic is of interest and primary outcome is a score, critical readers should seek answers to the following questions (Table 1). (If you can’t answer these questions, it’s tough to know what counts as an important effect.)

TABLE 1
QUESTIONS TO ANSWER TO UNDERSTAND A SCORE*

QUESTION	ANSWER
What’s being measured?	
Which end is up?	
What’s possible?	
What are some benchmarks?	
What matters?	

*Finding the answers can be challenging. One excellent resource for understanding functional health scores is McDowell I, Newell C. *Measuring Health, 2nd ed.* Oxford: Oxford Univ Pr; 1996.

What’s being measured?

The first step is to try to get a handle on the construct. This can be harder than you think. Like so many things in medicine, scores often go by their acronym (and even when you know what the acronym stands for, you may not be that much closer to the construct). Consider the following examples. PCS stands for physical component summary; it is an overall measure of physical function assessed by self-report (part of the Medical Outcomes Study SF-36). APACHE II stands for Acute Physiologic and Chronic Health Evaluation (second version); it is a prognostic measure for intensive care unit patients that is used to predict inpatient mortality.

Which end is up?

Sometimes it’s hard to know whether a higher score is a good thing or a bad thing. A high PCS score, for example, is good. A high APACHE II score, on the other hand, most definitely is not.

What’s possible?

Knowing the range of possible values is the next step for getting a feel for the results. Some scores, such as the PCS score, range from 0 to 100. But many do not (APACHE II ranges from 0 to 71).

What are some benchmarks?

The reader needs context—some grounding on what an expected score would be for a defined set of individuals. Published norms are available for the PCS score.¹ For example, in the general U.S. population, the average PCS score for men over 65 years is 42. A healthy 40-year-old will have an APACHE II score of 0.

What matters?

Finally, the reader needs help to make judgments about what constitutes an important change. In other words, a reader needs a clinical correlation. A 5-point decrease in the PCS score, for example, is equivalent to developing a new chronic disease like congestive heart failure. Of course, the information is not as precise as we would like (the severity of congestive heart failure varies from person to person, as does its impact), but it’s a lot better than nothing. A change in APACHE II from 12 to 24 is associated with an absolute increase in inpatient mortality of 30% (from approximately 10% to over 40%).

To make sense of scores, readers should try to answer the preceding questions. Unfortunately, authors often fail to provide the needed information. In these cases, if readers want to really understand what a score means, they must do the hard work themselves.

Reference

1. SF-36 Physical and Mental Health Summary Scales: A User’s Manual. Boston: The Health Institute, New England Medical Center; 1994.