Outcomes Part I
What Makes a Good Outcome Measure?

At last year’s Research in Gerontological Nursing’s annual editorial board meeting, members requested more information on how to choose high-quality outcomes for clinical trials (e.g., randomized controlled trials [RCTs]). As measurement quality has preoccupied a lot of my professional life, I thought I would briefly discuss what, in my opinion, makes a good outcome. In the next issue, I’ll discuss qualities of poor outcome measures. A plethora of detailed information and advice has been written about each of my points. I hope that the compilation of the eight qualities that follow provides a helpful list of factors to consider when designing or evaluating research outcomes and measurement.

Clinically Meaningful Outcomes
To engage in high-quality scientific pursuits in gerontological health care, we need to focus clinical trial research on clinically meaningful outcomes. If an intervention yields changes in measures such as biomarkers, but fails to make an older adult feel better, function better, or survive longer, it may not be a particularly useful intervention. Clinically meaningful outcomes address what is considered important by older adults. When deciding on what is a clinically meaningful outcome, researchers need to designate an amount of change that makes the difference clinically meaningful. How much of a decrease in night sweats, blood pressure, or nighttime awakenings is clinically important? Minimal clinically important difference (MCID) is the term used to indicate the smallest difference in a measurable clinical parameter that indicates a meaningful change in a health care outcome. It can be difficult to determine a valid parameter of meaningful change, particularly when biomarkers are outcomes. For a biomarker to qualify as a MCID, the change in score needs to relate to another measure, such as a change in needed medication, or a patient report of improvement. Patient reports of improvement can yield biased results, and it is difficult to determine their validity. Another approach is to determine statistically if the change is larger than what is expected by either the random variation of the sample or the measurement error of the instrument. If an outcome that is considered important by older adults yields a lot of measurement error, another outcome may in the long run produce more valid and useful findings. For example, although quality of life is considered a clinically meaningful outcome across most cancer types, a working group from the American Society of Clinical Oncology decided to measure symptoms rather than quality of life because of difficulty measuring and interpreting even validated quality of life measures (Ellis et al., 2014).

Sensitivity to Change
Clinical trials are trying to capture change. To detect the effect of an intervention, a measure should change in response to the treatment while remaining relatively unchanged if the treatment is not given. Methodologically, this is frequently discussed as capturing “signal” rather than “noise” or variability in the measure that is attributable to another source. Choose outcomes that are likely to change and do not have floor or ceiling effects. Stable constructs such as personality are not expected to change much or change easily. Cognitive status can be expected to change in delirium and depression but may be less likely to change as a result of an intervention for a person with Alzheimer’s disease. The demarcations or gradients of scoring also need to be fine enough to detect the changes you expect.

The Process of Change
Our overreliance on measuring means from one point in time does not allow us to understand the process of
change and may lead to a failure to detect an effect of the intervention that is present. We should not only want to understand the effectiveness of interventions, but also the process and time course of changes in outcomes. Analytically speaking, the process of change involves (a) the shape of the change, (b) the moderators of change, and (c) explanations for how the change occurs.

Measuring effects across a time-course plot shows the shape of change and can help explain the process of change. The duration and timing of effects can help us understand the timing and probability of relapse, as well dose-response relations. Outcome variables should be collected at a rate that reflects the dynamic nature of change resulting from, for example, a physical or behavioral intervention. A trajectory of change may be linear or non-linear. Change in outcomes may be rapid early in treatment, then stabilize, and then show another shift. Hierarchical linear modeling, growth curve analysis, and simple raw data graphing with multiple time points can help uncover these patterns.

Individual differences that affect treatment outcomes are commonly called moderator variables. Moderators help us understand differential effectiveness of the treatment for certain subgroups or conditions. Understanding moderators can help target interventions to those most likely to benefit, improve the quality of health care delivered, and save health care resources. Examples of moderator variables are social support, comorbid problems, and gender.

Often we infer mechanisms of action based on treatment effects and a theoretical framework, but this is not good enough. Treatment mechanisms allow us to understand how the intervention works to influence outcomes, and are exceedingly important for advancing theory, science, and the quality of health care delivered. Mediation is not the same as mechanism of action. To understand the mechanisms of action for a particular treatment, that mechanism must be measured along the time course of treatment effects. When treatments have multiple components and multiple causal mechanisms, the elucidation of mechanisms based on evidence can become quite thorny. To understand mechanisms of action, there must initially be a strong association between the intervention and measure of the proposed mechanism, as well as a relationship between the mechanism and outcome. The ability to demonstrate that there are not multiple causal paths strengthens the claim. Showing that a higher dose of the intervention increases the activation of the mediator and the effects also provides evidence for the mechanism of action. Replicating the results across samples, situations, and conditions, and demonstrating a logical explanation for the mechanism that is consistent with other scientific research, also aid acceptance of the credibility of the mechanism of action (Kazdin, 2007).

OFF-TARGET EFFECTS

When we conceptualize an intervention, we concentrate our efforts on measuring the outcomes to be achieved as well as sometimes measuring mediators and moderators. Consider measuring consequences that are different from the intended outcome and may contribute to unintended negative consequences. Given that interventions are designed to do something, even an ineffective intervention may be producing some off-target effects. People with dementia, for example, are susceptible to negative effects from too much environmental stimulation, and older adults in general have a smaller range between toxic and therapeutic doses of drugs. Measuring potential off-target mediators and outcomes may help move science forward in our understanding of effective interventions and mechanisms of action, and groups and subgroups most likely to benefit from an intervention.

INTERVENTION COSTS

Limits on health care resources mandate that costs of interventions are considered relative to benefits. Fiscal measures are increasingly expected in RCTs. Costs include measures of the interventionist’s time in planning and delivering the treatment, the assistive personnel’s time spent scheduling meetings, overhead costs, as well as any additional costs of treatment. There are many methodological approaches to cost-effectiveness analysis, and obtaining accurate data that will validly represent costs and benefits can be complex. Engaging health care economists on research teams is essential and also strategic for increasing the significance and quality of a research proposal.

SELF-REPORT AS A GOOD MEASUREMENT IDEA

Medical and health care research has been criticized for the passive role it assigns to patients in investigating their condition. Increased effort has been devoted to involving patients and advocacy groups in the design of studies as well as using patient reports to evaluate treatment outcomes. Clearly, there are times when a measure of people’s perceptions of their well-being, quality of life, functional ability, or pain level are more meaningful and clinically useful outcomes than objective measures. Other variables such as fatigue, emotional distress, attitudes, values, experiences, and beliefs are most directly assessed through self-report.
In 2004, the National Institutes of Health (NIH) launched an initiative to develop more measures of self-reported health. The resulting Patient-Reported Outcomes Measurement Information System (PROMIS) and Assessment Center is easily accessed (https://www.assessment-center.net/documents/InstrumentLibrary.pdf). A variety of tools for assessing physical, mental, and social health are available. One advantage of using these tools is that they may allow the researcher to better examine results across studies. A potential disadvantage in gerontology is that the tool may have been developed for all adults and thus may or may not be reliable or valid with older adult populations, or to measure specific geriatric outcomes.

FEASIBILITY
Feasibility is not a primary consideration, but it is necessary that the data can be obtained and that the data collection will not have unintended negative consequences on other aspects of the study. For example, a data collection method that induces stress when testing an intervention designed to reduce stress could confound results.

MARKETABLE OUTCOMES
Although the notion of choosing marketable outcomes may position science as a commodity, to conduct high-quality research we must seek and obtain funding. Each funding agency or organization has a certain set of priorities and valued outcomes. For example, the NIH directs its awards to research that improves the public’s health. Biobehavioral, functional, behavioral symptom management, self-management, and cost are outcomes that are commonly applicable to gerontological RCTs.

CONCLUSION
A lot of good measurement comes down to knowing what you want to know, and how to find it. There is a story of the legendary drunk who lost a coin and walked under a streetlight to find it. An observer said, “This is not where you lost it; so you won’t find it here.” If we are looking for an outcome that won’t be found using simple superficial measures, we must instead use measures that offer the potential of capturing real change as a result of an intervention.

REFERENCES

Christine R. Kovach, PhD, RN, FAAN, FGSA
Editor

The author has disclosed no potential conflicts of interest, financial or otherwise.

doi:10.3928/19404921-20191024-01