Defining the Primary Outcomes and Justifying Secondary Outcomes of a Study: Usually, the Fewer, the Better

Thomas R. Vetter, MD, MPH,* and Edward J. Mascha, PhD†

One of the first steps in designing and conducting a research study is identifying the primary and any secondary study outcomes. In an experimental, quasi-experimental, or analytic observational research study, the primary study outcomes arise from and align directly with the primary study aim or objective. Likewise, any secondary study outcomes arise from and directly align with any secondary study aim or objective. One designated primary study outcome then forms the basis for and is incorporated literally into the stated hypothesis. In a Methods section, authors clearly state and define each primary and any secondary study outcome variable. In the same Methods section, authors clearly describe how all primary and any secondary study outcome variables were measured. Enough detail is provided so that a clinician, statistician, or informatician can know exactly what is being measured and that other investigators could duplicate the measurements in their research venue. The authors provide published substantiation (preferably) or other documented evidence of the validity and reliability of any applied measurement instrument, tool, or scale. A common pitfall—and often fatal study design flaw—is the application of a newly created (“home-grown”) or ad hoc modification of an existing measurement instrument, tool, or scale—without any supporting evidence of its validity and reliability. An optimal primary outcome is the one for which there is the most existing or plausible evidence of being associated with the exposure of interest or intervention. Including too many primary outcomes can (a) lead to an unfocused research question and study and (b) present problems with interpretation if the treatment effect differed across the outcomes. Inclusion of secondary variables in the study design and the resulting manuscript needs to be justified. Secondary outcomes are particularly helpful if they lend supporting evidence for the primary endpoint. A composite endpoint is an endpoint consisting of several outcome variables that are typically correlated with each. In designing a study, researchers limit components of a composite endpoint to variables on which the intervention of interest would most plausibly have an effect, and optimally with preliminary evidence of an effect. Ideally, components of a strong composite endpoint have similar treatment effect, frequency, and severity—with the most important being similar severity. (Anesth Analg 2017;125:678–81)

Identifying and Defining the Primary and Secondary Study Outcomes

Study Aim/Study Outcome/Study Hypothesis

One of the first steps in designing and conducting a research study is identifying the primary and any secondary study outcomes. This key step of choosing the most appropriate outcome variables to address the study aims or objectives is applicable with any experimental, quasi-experimental, or analytic observational research design.

In an experimental, quasi-experimental, or analytic observational research study, the primary study outcomes logically arise from and align directly with the primary study aim or objective. One designated primary study outcome then forms the basis for and is incorporated literally into the stated primary null hypothesis or alternative hypothesis.

In a similar fashion, each secondary study outcome logically arises from and directly aligns with a secondary study aim or objective. A designated secondary study outcome then forms the basis for and is incorporated literally into the respective stated secondary null hypothesis or alternative hypothesis.

State and Define All Outcome Variables in the Methods Section

In a labeled subsection within the Methods section (“Outcome Variables”), the authors clearly state and define
each primary and any secondary study outcome variable. This information can be summarized in the text and also presented in a separate table if deemed to be more reader-friendly and space efficient, particularly if there are detailed definitions that accompany the outcome measures. Authors also at least state the primary outcome variables in the Background or Methods section of the manuscript Abstract.

Authors should not introduce or report on any additional outcome variables in the Statistical Methods, Results, or Discussion or in the tables/figures of their manuscript. Of course, additional outcome variables can be included in their description of proposed future research in the Discussion.

**Describe How the Outcome Variables Were Measured in the Methods Section**

In the same labeled subsection within the Methods section (“Outcome Variables”), authors clearly describe how all the primary and any secondary study outcome variables were measured. Enough detail is provided so that a clinician, statistician, or informatician can know exactly what is being measured and so that other investigators could duplicate the measurements in their research venue. This information can be included in the above designated table. Alternatively, a more complex or diagrammatic measurement instrument, tool, or scale can be presented in its entirety as a separate figure (with appropriate creator or copyright attribution).

The authors provide published substantiation (preferably) or other documented evidence of the validity and reliability of any applied measurement instrument, tool, or scale:1–3 A common pitfall—and often fatal study design flaw—is the application of a newly created (“home-grown”) or ad hoc modification of an existing measurement instrument, tool, or scale—without any supporting evidence of its validity and reliability.

**Provide the Scoring Range for All Outcome Measures**

Whenever applicable, authors provide the utilized scoring range for each of their study outcome measures. If the utilized scoring range has been achieved by collapsing or stratifying (dividing) an underlying continuous variable into separate categories (eg, yes/no; low/moderate/high), a cogent rationale and any previously published support are provided. Substantial information and statistical power are often lost by such categorization, so it should be avoided unless a strong clinical or statistical rationale can be given.

**Provide the Minimal Clinically Important Difference**

The “minimally important difference” is an innately appealing concept as one attempts to compare and to interpret patient-reported health outcomes.4

The term minimal clinically important difference (MCID) was first described by Jaeschke et al5 in 1989, who insightfully noted that although a statistically significant change can occur—based on a health-related instrument (assessment tool) that measures health status before versus after a medical or surgical intervention—in some instances, the observed change is not clinically important.6 Likewise, the observed change or difference in an outcome of interest may exceed the accepted MCID, yet not achieve statistical significance.

Jaeschke et al5 originally defined the MCID as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.” The MCID is thus innately patient centered as it represents the smallest amount an outcome must change to be meaningful to patients.5,7

“There are many faces to the MCID, it is not a simple concept, nor simple to calculate.”8 Nevertheless, whenever applicable and available, authors provide the MCID for their primary study outcomes in the Methods section of their manuscript. The MCID for a measurement instrument (assessment tool) may have been previously reported when it was first developed or its use first reported. However, the MCID may not have yet been documented, in which case the current investigators need to estimate it based on previously reported, similar outcomes data.

Once identified or estimated for the primary outcome, the MCID serves two purposes for a study. It can be used in an a priori power analysis to estimate the appropriate sample size and also to interpret results once the study is completed. For example, a negative study (nonsignificant P value) for which the 95% confidence interval around the treatment effect does not contain the MCID supports a definitive negative conclusion. However, if the confidence interval contains the MCID, the negative conclusion would not be definitive.

**Specific Examples of Published Studies**

Benyamin and Staats,9 Staats and Benyamin,10 and the other Mild Decompression Alternative to Open Surgery Evidence-based Neurogenic Claudication Outcomes Research Investigators clearly state the objective of their randomized controlled trial was to compare patient’s functional disability, claudication symptoms, and pain intensity, following treatment with either minimally invasive lumbar decompression (Mild; Vertos Medical, Aliso Viejo, CA) or epidural steroid injection in lumbar spinal stenosis patients exhibiting neurogenic claudication and having verified ligamentum flavum hypertrophy as a contributing factor. Benyamin and Staats9 and Staats and Benyamin10 then clearly state that their corresponding primary outcome measure was the Oswestry Disability Index and secondary outcomes were the Zurich Claudication Questionnaire and numeric pain rating scale. These authors provide ample published evidence of the utility of these 3 chosen outcome measures, as well as their respective scoring range and MCID (accepted “minimal important change”). They defined clinical responders as patients experiencing the 10-point improvement (accepted “minimal important change”) in the Oswestry Disability Index at follow-up and used this binary outcome in their a priori power analysis and minimal sample size determination.9,10

Sultan et al11 clearly state the objective of their meta-analysis was to determine whether low- or high-dose intrathecal morphine provides acceptable duration and intensity of analgesia with fewer side effects after elective cesarean delivery.
Sultan et al\textsuperscript{11} then clearly state that their primary outcome was the duration of analgesia, defined as the elapsed time from intrathecal morphine administration until the patient’s first request for analgesia. Their clearly stated secondary outcomes include other measures of analgesia (pain intensity scores and morphine use), maternal side effects (nausea, vomiting, pruritus, and respiratory depression), and neonatal outcomes (umbilical arterial and venous pH and Apgar scores at 1 and 5 minutes). All of these secondary outcomes, and any applied derivations, are clearly defined.\textsuperscript{11}

**LIMITING THE NUMBER OF PRIMARY STUDY OUTCOMES**

An optimal primary outcome is the one for which there is the most existing or plausible evidence of being associated with the exposure of interest or intervention. In a previous statistical tutorial, we discussed the merits of limiting the number of research study hypotheses.\textsuperscript{12} Doing so will naturally limit the number of primary outcomes. However, for a given study aim and hypothesis, there are often several outcome variables that researchers consider as related to their exposure or intervention of interest. For example, they might hypothesize that an intraoperative intervention (eg, spinal or regional anesthesia versus general anesthesia) reduces serious infection, opioid consumption, readmission to the hospital, and 30-day mortality. Including all these as primary outcomes would (a) lead to an unfocused research question and study and (b) present problems with interpretation if the treatment effect differed across the outcomes. Having multiple primary outcomes should also be avoided if at all possible because it increases the chance of reporting a false positive finding.

**JUSTIFYING ANY ADDITIONAL, SECONDARY STUDY OUTCOMES**

Inclusion of secondary outcomes in the study design and the resulting manuscript needs to be justified. Secondary outcomes are particularly helpful if they lend supporting evidence for the primary endpoint. In the above example, investigators might choose 30-day mortality as the primary endpoint but include the other 3 (infection, readmission, and opioid consumption) as secondary ones. They might argue that an observed effect (or lack of effect) on mortality would be interpreted more easily if differences were also found (or not found) on the other 3 variables. In such a scenario, secondary variables might even be considered as mechanism or mediator variables between the intervention and primary outcome and analyzed accordingly.\textsuperscript{13} Thus, secondary outcomes can help tell the whole story.

However, we caution against the not-so-rare tendency to include a myriad of secondary outcome variables, whose results are reported alongside the primary outcome variable—with little distinction between primary and secondary outcomes throughout the manuscript, including in the Discussion and Conclusions sections. Such manuscripts can be quite unfocused.

**THE USE OF A COMPOSITE STUDY ENDPOINT**

A composite endpoint is an endpoint consisting of several outcome variables that are typically correlated with each other. For example, a composite study endpoint might consist of several binary events (eg, perioperative myocardial infarction, stroke, mortality), which are combined as a single all-or-none event (a “collapsed composite”) or alternatively are analyzed as individual events while estimating an overall treatment effect across them.\textsuperscript{14,15} As noted by Mascha and Sessler,\textsuperscript{14} a composite outcome is frequently chosen as the primary endpoint because (a) no single outcome fully characterizes the disease or overall outcome of interest (eg, a major adverse cardiac event) and/or (b) the individual outcomes are rare and thus the statistical power for even a very large sample would be inadequate to demonstrate a significant association or effect for any single one. In designing a study, researchers limit components of a composite endpoint to variables on which the intervention of interest would most plausibly have an effect, and optimally with preliminary evidence of an effect. The tendency to include any variable with a remotely possible relationship to the exposure will likely lead to a negative study. This is the case because, eg, if there is no effect for most of the components (say 4 of 5), then even though the fifth component is markedly affected by the intervention, the “zero” effects will drive the overall (composite) result to be clinically and statistically nonsignificant and “wash out” the true effect. Ideally, components of a strong composite endpoint have similar treatment effect, frequency, and severity—with the most important being similar severity.

While composite outcomes can help capture a complex disease manifestation of interest and may be associated with increased statistical efficiency (power), they are also fraught with important potential problems and should be used judiciously.\textsuperscript{14,15} For example, reporting only the overall treatment effect on the composite outcome can mask substantial treatment effect heterogeneity among the individual components of the composite. Therefore, in such studies, the statistical plan always assesses the treatment-by-component interaction, either statistically or at least descriptively.

**Specific Examples of Published Studies**

The perioperative ischemic evaluation study trial\textsuperscript{16} did observe a statistically and clinically significant treatment-by-component interaction, such that the β-blocker was associated with decreased myocardial infarction and was also associated with increased stroke and mortality. Main results of the study thus appropriately focused on its individual components and not the composite.

Alternatively, the National Institute of Neurological Disorders and Stroke multicenter trial on intravenous recombinant tissue plasminogen activator for ischemic stroke\textsuperscript{17} hypothesized that the drug would be beneficial for neurological function if and only if it reduced each of 4 components of a composite outcome, including binary versions of the National Institutes of Health Stroke Scale, Barthel Index, Rankin Scale, and Glasgow Coma Score. All components were improved, as well as the composite endpoint.

Last, the observational study by Bardia et al\textsuperscript{18} assessed the association between two exposures, preoperative hemoglobin A1C and postoperative glycemic variability, and a composite outcome of 30-day major adverse outcomes after cardiac valve surgery. The authors included components that have clear potential for being affected by the exposures. An issue is that the frequencies and severities of
the outcome components vary considerably, suggesting that their choice of components might have been more carefully considered.

CONCLUSIONS
In choosing primary and any secondary study outcomes, a judicious approach should be taken, in which fewer is usually better. Secondary outcomes especially need to be justified, lest a “laundry list” be measured and data-mining seemingly occurs. Likewise, while composite outcomes can help capture a complex disease manifestation of interest and may be associated with increased statistical efficiency (power), they are also fraught with important potential problems and should be used judiciously. Authors should always clearly state and define each primary and any secondary study outcome variable and how they were measured.

DISCLOSURES
Name: Thomas R. Vetter, MD, MPH.
Contribution: This author helped write and revise the manuscript.
Name: Edward J. Mascha, PhD.
Contribution: This author helped write and revise the manuscript.
This manuscript was handled by: Jean-Francois Pittet, MD.

REFERENCES