ELSEVIER

Journal of Clinical Epidemiology ■ (2023) 111242

Journal of Clinical Epidemiology

KEY CONCEPTS IN CLINICAL EPIDEMIOLOGY

Surrogate endpoints: a key concept in clinical epidemiology

Robin Christensen[a,b,*], Oriana Ciani[c], Anthony M. Manyara[d,e], Rod S. Taylor[f,g]

[a]*Section for Biostatistics and Evidence-Based Research, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark*
[b]*Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark*
[c]*Centre for Research on Health and Social Care Management, SDA Bocconi School of Management, Milan, Italy*
[d]*MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University of Glasgow, Glasgow, UK*
[e]*Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK*
[f]*Robertson Centre for Biostatistics, School of Health and Well Being, University of Glasgow, Glasgow, UK*
[g]*Faculty of Health Sciences, National Institute of Public Health and Department of Psychology, University of South Denmark, Odense, Denmark*

Accepted 18 December 2023; Published online xxxx

Abstract

Surrogate endpoints are biomarkers or intermediate outcomes that are used as substitutes for clinical outcomes of interest, often to expedite research or decision-making. In contrast, patient-important (or patient-centered) outcomes are health outcomes that are of direct relevance and importance to patients themselves; clinical trials may have measured the impact of the intervention on other endpoints related to, but different from, those of primary importance to patients. This article aims to elaborate on the use and understanding of surrogate endpoints. There should be a well-understood and scientifically grounded relationship between the surrogate (replacement) and the patient-important (target) endpoint it is intended to represent. It should be biologically plausible that changes in the surrogate will consistently and predictably reflect changes in the patient-important endpoint. The surrogate endpoint should show a threshold effect, meaning that a specific change (or state) in the surrogate with an intervention (relative to the comparator) is associated with a predictable (change in the) patient-important outcome. This helps establish a meaningful cutoff or target for the treatment effect on the surrogate endpoint. While surrogate endpoints offer advantages in certain situations, it is important to remember that their use requires careful validation to ensure they reliably predict the true clinical outcome. The validity of ''surrogate endpoints'' should be supported by robust scientific evidence and rigorous evaluation before these can be considered and labeled as surrogate endpoints.

Keywords: Surrogate outcomes; Randomized controlled trials; Clinical epidemiology; Indirectness; Outcome measures; Endpoints

1. Background

Contemporary research emphasizes the perspective and experiences of the patient themselves, focusing on outcomes that are meaningful and important to individuals in terms of their health and daily life. Clinical evidence should preferably come from research that directly compares the interventions of interest, applied to the relevant population(s), and should measure outcomes important to patients, clinicians, and policymakers to be useful for decision-making [1,2]. In various contexts, the terms "outcome" and "endpoint" are used interchangeably. While an outcome broadly refers to any result or consequence of a particular action (e.g., intervention) in a clinical study, an endpoint refers to a specific measurable variable used to assess the success or effectiveness of a clinical trial. As an example, the choice of primary endpoint when designing a trial is crucial; it serves as the primary focus of the study's sample size, analysis, and interpretation. The results related to the primary endpoint are pivotal in the assessment of whether the treatment effect compared to the control is both statistically significant and clinically meaningful. The distinction between the two terms ("outcome" and "endpoint") may vary across disciplines, so it may be important to consider the specific context in which they are being used.

While patient-important (or patient-centered) outcomes are health outcomes that are of direct relevance and importance to patients themselves, clinical trials may have

Funding: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL).

* Corresponding author. Section for Biostatistics and Evidence-Based Research, Department of Clinical Research, The Parker Institute, Nordre Fasanvej 57, DK-2000 Copenhagen F, Denmark. Tel.: +45-3816-4165; fax: +45-3816-4159.

E-mail address: Robin.Christensen@regionh.dk (R. Christensen).

https://doi.org/10.1016/j.jclinepi.2023.111242

measured the impact of the intervention on other endpoints related to, but different from, those of primary importance to patients. Examples include the use of biomarkers and intermediate outcomes to assess the apparent efficacy or effectiveness instead of applying a directly applicable patient-important outcome [3]. When used for this purpose, biomarkers and intermediate outcomes potentially become surrogate endpoints. Surrogate endpoints can be defined as successful biomarkers or intermediate outcomes that are used as substitutes for clinical outcomes of interest, often to expedite research or decision-making. Thus, surrogate endpoints are not themselves necessarily important to patients but rather act as substitutes for the ''real deal'' in the presumption that changes in the surrogate also reflect changes in an outcome important to patients [4]. While surrogate endpoints might appear degraded as they contrast with the ''what is really important,'' they are frequently used in medical research and clinical trials when the patient-important outcome may be impractical, time-consuming, costly, or ethically challenging [5].

2. Validity of a surrogate endpoint

Attempts to systematically assess definitions and success of surrogate endpoints in trials have been narrative in nature and restricted to specific clinical/health areas; there is no published comprehensive guidance for reporting surrogate endpoints as primary outcomes [5]. The credibility of a surrogate endpoint in research refers to its ability to predict or substitute for a clinical outcome of interest. The credibility depends on whether it meets certain criteria. The validity of a surrogate endpoint is often assessed based on variations of the following characteristics [4,6]:

- *Biologic plausibility:* There should be a well-understood and scientifically grounded relationship between the surrogate (replacement) and the clinical (target) endpoint it is intended to substitute. It should be biologically plausible that changes in the surrogate will consistently and predictably reflect changes in the clinical endpoint.
- *Epidemiologic evidence:* Empirical evidence from observational studies or previous clinical trials should support the relationship between the surrogate (substitute) and the clinical (patient-important) outcome. Multiple studies conducted across different populations or settings should consistently demonstrate a strong association between the surrogate (the independent variable [X]) and clinical (dependent variable [Y]) outcome. If the intervention leads to changes in the surrogate endpoint and these changes are consistently associated with changes in the clinical outcome, it provides evidence of a causal relationship between the intervention, the surrogate endpoint, and the clinical outcome.
- *Analytical validation:* Statistical analyses should be conducted to assess the strength and consistency of the relationship between the surrogate endpoint and the clinical outcome. These analyses include correlation coefficients, regression models, or other statistical methods to quantify the association. Establishing causal inference between a surrogate endpoint (X) and a clinical outcome (Y) involves a combination of experimental design, statistical analysis, and a solid understanding of the underlying biological mechanisms.
- *Surrogate threshold effect:* The surrogate endpoint should show a threshold effect, meaning that a specific change in the surrogate is associated with a predictable change (or state) in the clinical outcome. This helps establish a meaningful cutoff or target for the surrogate endpoint.

Although not explicitly mentioned, the Bradford Hill criteria (for causation) can be relevant in evaluating the strength of evidence for a surrogate endpoint. While consistency and specificity may not be explicitly named criteria for evaluating surrogate endpoints, other aspects such as validity, reliability, and responsiveness are commonly considered. The focus is on understanding the ability of the surrogate endpoint to accurately represent changes in the true clinical outcome and whether this relationship holds across different scenarios. For example, criteria such as biological plausibility, consistency, and strength of association are directly applicable when evaluating whether changes in a surrogate endpoint reliably predict changes in a clinical outcome. Ensuring that surrogate markers are sensitive to treatment effects obviously enhances their utility as valid substitutes for true clinical outcomes in various research and clinical settings. If a surrogate endpoint satisfies all these criteria, it may be considered a valid and useful substitute in clinical trials. If, on the other hand, the surrogate endpoint lacks a clear biological basis or shows inconsistency in its association with the clinical outcome, its validity may be questioned.

Rigorous and systematic evaluation of proposed surrogate endpoints is crucial for making informed decisions about the efficacy and safety of interventions in medical research and practice. Just like in any causal inference scenario, researchers need to consider potential confounding variables that could affect both the surrogate and the clinical outcome. If there are uncontrolled confounders, the observed relationship between the surrogate (X) and the clinical outcome (Y) might not be solely due to the intervention but could be influenced by these other confounding factors. Accordingly, evidence should preferably be synthesized (e.g., via meta-regression analysis) addressing the question if the evidence from randomized trials consistently shows that improvement in the surrogate endpoint has consistently led to improvement in the (patient-important) target outcome.

3. Example

A frequently used example of a successful surrogate endpoint is in the field of cardiovascular disease. In this context, low-density lipoprotein (LDL) cholesterol is a biomarker used as a surrogate endpoint for cardiovascular morbidity when assessing the efficacy of cholesterol-lowering medications (e.g., statin therapy). Biological plausibility: There is a well-established understanding of the relationship between elevated LDL cholesterol levels and the risk of cardiovascular events, such as heart attacks and strokes (i.e., lipids and lipoproteins are involved in atherosclerosis). Lowering LDL cholesterol levels is known to reduce the risk of these events. Epidemiological studies have consistently shown that lowering LDL cholesterol levels leads to a reduction in the risk of cardiovascular events. Changes in LDL cholesterol levels predict changes in the occurrence of clinical endpoints like heart attacks and strokes. Multiple clinical trials and observational studies across diverse populations have demonstrated that the reduction in LDL cholesterol levels achieved on different pharmacological agents (usually) translates into a lower risk of cardiovascular events. Over the years, numerous clinical trials have validated the use of LDL cholesterol reduction as a predictor of reduced cardiovascular risk. Trials have shown that a successful reduction in LDL cholesterol levels (e.g., 1 mmol/L) is directly associated with significant reductions in cardiovascular events. Cardiovascular events are critical clinical outcomes with substantial implications for patients' health and quality of life; reducing the risk of heart attacks and strokes is a meaningful clinical goal.

While direct assessment of cardiovascular events requires long-term follow-up and large sample sizes, measuring LDL cholesterol levels is a more feasible and shorter-term approach to assessing treatment effects [7]. LDL cholesterol reduction to prevent major adverse cardiovascular events (MACE) varies depending on the patient population, baseline risk factors, and the specific guidelines being followed. For instance, individuals with moderate risk factors for cardiovascular disease might be advised to achieve an LDL cholesterol level between 70 and 100 mg/dL (1.8−2.6 mmol/L). Clinical trials evaluating statins have reported relative risk reductions in the range of approximately 20−50% for MACE. Although these figures are controversial, this means that, on average, individuals taking statins could have up to 50% lower risk of experiencing a cardiovascular event compared to those not taking statins.

4. Pointers

Surrogate endpoints are biomarkers or intermediate outcomes that are used as substitutes for clinical outcomes of interest, often to expedite research or decision-making. As depicted in Table 1, the use of surrogate endpoints (both in individual trials and in decision-making) requires careful consideration to ensure that they reliably predict the desired clinical outcome (i.e., the patient-important outcome cannot just be implied). The scientific and medical community continuously evaluates and refines the use of surrogate endpoints based on accumulating evidence and experience. Regulatory agencies such as the US Food and Drug Administration should be encouraged to assess the validity of surrogate endpoints according to standardized frameworks to determine their acceptance for evaluating the efficacy and safety of new medical interventions.

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group is a widely recognized international collaboration of researchers and methodologists who have developed a framework for evaluating the quality of evidence and making recommendations in healthcare via rigorous systematic reviews, clinical guidelines, and health technology assessments. The GRADE Working Group has provided pragmatic guidance on the appropriate use of surrogate outcomes in evidence synthesis and guideline development. The GRADE approach recognizes that surrogate outcomes may be seen as a necessity and thus employed as major outcomes or maybe even as the primary endpoint in individual trials or as intermediate outcomes in a causal pathway between the intervention and the patient-relevant clinical outcomes [8].

While surrogate endpoints offer advantages in certain situations, it is important to remember that their use

Table 1. Points to consider when interpreting evidence from a surrogate endpoint.

Indirectness
If the relationship between the surrogate and the clinical (patient-important) outcome is not well-established or is uncertain, it can lead to indirectness in the evidence (less certainty with the apparent findings).
Lack of validation
If there is a lack of analytical validation studies demonstrating the credibility of the surrogate endpoint (X) in predicting clinical outcomes (Y), the quality of evidence might be rated down due to uncertainty.
Magnitude of effect
If the treatment effect on the surrogate endpoint is substantial but the effect on the clinical outcome is uncertain or smaller, this will reduce our certainty in the apparent findings.
Risk benefit balance
If using the surrogate endpoint as a basis for decision-making could lead to potential harm or inappropriate treatment decisions, a guideline panel is more likely to provide a conditional (weak) recommendation.
Uncertainty
The use of surrogate endpoints will introduce additional uncertainty into the evidence. If there is significant uncertainty regarding the relationship between the surrogate and clinical outcomes, the quality of evidence might be rated as less credible.

requires careful validation to ensure they reliably predict the true clinical outcome. The use of biomarkers or intermediate outcomes as a replacement for clinical endpoints should be supported by robust scientific evidence and rigorous evaluation before a claim of surrogate endpoints can be made.

CRediT authorship contribution statement

Robin Christensen: Conceptualization, Funding acquisition, Methodology, Project administration, Writing − original draft, Writing − review & editing. **Oriana Ciani:** Conceptualization, Writing − original draft, Writing − review & editing. **Anthony M. Manyara:** Conceptualization, Writing − original draft, Writing − review & editing. **Rod S. Taylor:** Conceptualization, Methodology, Supervision, Writing − original draft, Writing − review & editing.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

[1] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395−400.

[2] Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Res Synth Methods 2013;4(1):49−62.

[3] Maksymowych WP, Fitzgerald O, Wells GA, Gladman DD, Landewé R, Ostergaard M, et al. Proposal for levels of evidence schema for validation of a soluble biomarker reflecting damage endpoints in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis, and recommendations for study design. J Rheumatol 2009;36:1792−9.

[4] Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. JAMA 1999;282:771−8.

[5] Manyara AM, Davies P, Stewart D, Weir CJ, Young AE, Wells V, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. J Clin Epidemiol 2023;160:83−99.

[6] Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008;17(3):303−40.

[7] Buhr KA. Surrogate end points in secondary analyses of cardiovascular trials. Prog Cardiovasc Dis 2012;54(4):343−50.

[8] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence–indirectness.; GRADE Working Group. J Clin Epidemiol 2011;64:1303−10.

Further reading

[1] Stamp LK, Frampton C, Morillon MB, Taylor WJ, Dalbeth N, Singh JA, et al. Association between serum urate and flares in people with gout and evidence for surrogate status: a secondary analysis of two randomised controlled trials. Lancet Rheumatol 2022;4:E53−60.

An empirical validation of the threshold effect of getting serum urate below 6 mg/dL (0·36 mmol/L) in patients with gout to reduce the risk of having a gout flare.

[2] Available at: https://wwwgradeworkinggroup.org/. Accessed February 3, 2024.

Link to the GRADE Working Groups website, incl. materials on indirectness.

[3] Manyara AM, Davies P, Stewart D, Weir CJ, Young AE, Wells V, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. J Clin Epidemiol 2023;160:83−99.

A scoping review from the SPIRIT|CONSORT group on definitions of surrogate outcomes.

[4] Manyara AM, Davies P, Stewart D, Weir CJ, Young A, Butcher NJ, et al. Protocol for the development of SPIRIT and CONSORT extensions for randomised controlled trials with surrogate primary endpoints: SPIRIT-SURROGATE and CONSORT-SURROGATE. BMJ Open 2022;12:e064304.

Protocol for the development of SPIRIT and CONSORT extension for reporting on surrogate outcomes.