

Surrogate Endpoints in Clinical Research

Mariana A. Burgos¹ , Silvana B. Soli² , Habik A. Páez Álvarez²  and María F. Grande Ratti¹ 

¹Internal Medicine Research Area, Department of Internal Medicine, Hospital Italiano. Buenos Aires, Argentina

²Research Department, Universidad Hospital Italiano. Buenos Aires, Argentina

ABSTRACT

A surrogate endpoint is a biomarker or intermediate endpoint used as a substitute for a clinically relevant endpoint, such as survival or quality of life, and proposed to predict the benefits or harms of an intervention. Its use in clinical research may reduce the duration and sample size of clinical trials and expedite regulatory assessment and coverage decisions for new treatments. However, not every biomarker or intermediate endpoint constitutes a validated surrogate endpoint. Their use remains controversial because, in many cases, they have not been adequately validated as reliable predictors of clinically meaningful benefits for patients. This statistical note provides a critical review of the use of surrogate endpoints, examines their main conceptual and methodological limitations, and proposes an approach that prioritizes clinically relevant, patient-centered endpoints.

Keywords: surrogate outcomes, clinical research, biomarkers, clinical endpoints, health decision-making.

Desenlaces subrogados en investigación clínica

RESUMEN

Un desenlace subrogado es un biomarcador o desenlace intermedio usado como sustituto de un desenlace clínico relevante, como la supervivencia o la calidad de vida, propuesto para predecir los beneficios o daños de una intervención. Su utilización en investigación clínica puede reducir la duración y el tamaño de los ensayos clínicos, acelerar la evaluación regulatoria y las decisiones de cobertura de nuevos tratamientos. Sin embargo, no todo biomarcador ni todo desenlace intermedio constituye un subrogado validado. Su uso es objeto de debate, ya que en muchos casos, no han sido validados adecuadamente como predictores fiables de beneficios clínicos significativos para los pacientes. Esta nota estadística ofrece una revisión crítica del uso de desenlaces subrogados, analiza sus principales limitaciones conceptuales y metodológicas, y propone un enfoque que priorice desenlaces clínicamente relevantes y centrados en el paciente.

Palabras clave: desenlaces subrogados, investigación clínica, biomarcadores, desenlaces clínicos, toma de decisiones en salud.

Corresponding author: mariana.burgos@hospitalitaliano.org.ar, Burgos MA.

Received: 12/30/2025 | Accepted: 08/14/2026 | Published: 09/25/2026

DOI: <http://doi.org/10.51987/Rev.Hosp.Ital.B.Aires.v46i3.1312>

How to cite: Burgos MA, Soli SB, Páez Álvarez HA, Grande Ratti MF. Surrogate Endpoints in Clinical Research. *Rev HospItal B.Aires.* 2026;46(3):e0001312

INTRODUCTION

In the late 2000s, bevacizumab (a monoclonal antibody) was introduced for the treatment of metastatic breast cancer after several studies showed a significant improvement in progression-free survival, which was used as a surrogate outcome. These findings led the US Food and Drug Administration (FDA) to grant accelerated approval in 2008. However, subsequent studies confirmed that this benefit did not translate into improved overall survival and was associated with substantial toxicity, prompting the FDA to withdraw the indication in 2011.¹

In 2006, torcetrapib (an experimental drug developed to treat cardiovascular disease) emerged as a way to modify the lipid profile, producing marked increases in HDL cholesterol and significant reductions in LDL cholesterol. However, the risk of cardiovascular events and mortality increased significantly. The trial was terminated early, and torcetrapib was never approved, making it a clear example of how a favorable change in a surrogate outcome does not necessarily translate into actual clinical benefit.²

These cases show that surrogate outcomes do not always correspond to clinical benefits for patients. The literature provides numerous examples illustrating this discrepancy (Table 1). In this context, this Statistical Note examines the role of surrogate outcomes by reviewing their definition, uses, and limitations in everyday practice.

CONCEPT AND DEFINITION

The US Food and Drug Administration (FDA) and the National Institutes of Health (NIH) propose the following conceptual definitions:^{3,4}

- **Biomarker:** A characteristic that is objectively measured and evaluated as an indicator of biological processes, pathological processes, or pharmacological responses to a therapeutic intervention.
- **Intermediate outcome:** A variable that may be used as an outcome in a study but has not necessarily been validated as a substitute for a clinically relevant outcome.
- **Surrogate outcome:** A biomarker or intermediate outcome whose ability to substitute for a clinically relevant outcome and predict the benefit or harm of an intervention has been demonstrated through a formal validation process.⁵
- **Clinical outcome:** A characteristic or variable that reflects how a patient feels, functions, or survives.

According to Lassere, biomarkers and clinical outcomes are conceptually distinct variables that fall along a continuum from disease-centered to patient-centered measures.⁶

A biomarker is a variable with no immediate clinical meaning, whose interpretive value depends on an understanding of the underlying biological mechanisms and on accumulated clinical and epidemiological evidence. Consequently, it has no intrinsic validity and requires formal validation before it can be used as a substitute for a clinically relevant outcome.

In contrast, patient-centered variables have direct clinical relevance, are meaningful outcomes in their own

right, require no additional validation, and represent the standard for clinical decision-making and for use as primary outcomes in therapeutic trials.

An intermediate outcome is a variable within the causal pathway of a disease but has not been formally validated as a substitute for a clinical outcome. A surrogate outcome is defined as a biomarker or intermediate outcome that has undergone a rigorous validation process supporting its use as a substitute for a clinically relevant outcome⁶ (Figure 1).

Without a robust validation process, surrogate outcomes can lead to what Montori and Yudkin termed “surrogate idolatry”: indicators that take on a life of their own and drive clinical practice guideline recommendations and regulatory decisions but do not necessarily translate into clinically meaningful benefits for patients.⁷

USES AND APPLICATIONS IN PRACTICE

Measuring patient-relevant clinical outcomes, such as quality of life or mortality, requires prolonged follow-up and large sample sizes and, consequently, a considerable investment of time and resources. Surrogate outcomes allow interventions to be evaluated more efficiently and are particularly useful in rare diseases, pediatric populations, or pregnancy, where low incidence makes it difficult to recruit sufficient numbers of patients; in the development of innovative therapies, which require early evidence of biological efficacy; or during exploratory phases of research.^{5,8} Among the best-established examples, reductions in blood pressure have been shown to predict a decrease in cardiovascular events, particularly stroke, across multiple trials and different classes of antihypertensive drugs. Reductions in LDL cholesterol have been associated with a decrease in major cardiovascular events with statins. These examples show that surrogate outcomes can provide reliable estimates of clinical benefit when supported by causal plausibility and consistent evidence from multiple trials.

The problem arises when their use ceases to be the exception and becomes the predominant practice.

LIMITATIONS AND CONTROVERSIES

The indiscriminate use of surrogate outcomes raises methodological and ethical concerns. The evidence supporting them may be indirect: a pathophysiological association between a biomarker or intermediate outcome and the final outcome may be plausible but not necessarily predictive.⁶ This reduces the overall quality of the available evidence according to the GRADE system.

(Grading of Recommendations, Assessment, Development and Evaluation) system by lowering the certainty of evidence due to indirectness. In addition, a statistically significant improvement in a surrogate outcome may not translate into clinical benefit, may generate false-positive findings, or may even be associated with increased mortality or adverse events.⁹

The validity of a surrogate outcome is not a universal property of the marker; rather, it must be established

Table 1. Documented examples of the use of surrogate outcomes in the history of clinical research

Clinical condition	Surrogate outcome	Clinical outcome	Theoretical rationale	Reference studies	Observed evidence
Diabetes mellitus	HbA1c	Reduction in microvascular/macrovascular complications	Greater reductions in HbA1c would result in fewer complications.	ACCORD study ¹⁵ : 10,251 patients with T2DM at high cardiovascular risk. Intervention: intensive therapy (HbA1c < 6%) vs. standard therapy (HbA1c 7.0–7.9%). Results: HbA1c, 6.4% (intensive-therapy group) vs. 7.5% (standard-therapy group); ↓ cardiovascular events (HR, 0.90; 95% CI, 0.78–1.04; P = 0.16); ↑ all-cause mortality in the intensive-therapy group (HR, 1.22; 95% CI, 1.01–1.46; P = 0.04). ¹⁵	The use of intensive therapy to achieve normal HbA1c levels over 3.5 years increased mortality and did not significantly reduce major cardiovascular events. These findings identified a previously unrecognized harm of intensive glucose lowering in high-risk patients with T2DM.
Dyslipidemia	HDL LDL	Reduction in mortality Reduction in cardiovascular events Reduction in mortality	High-density lipoprotein (HDL) cholesterol levels are a strong and independent inverse predictor of cardiovascular disease.	ILLUMINATE study ² : 15,067 patients at high cardiovascular risk. Intervention: atorvastatin plus torcetrapib vs. atorvastatin alone. Results: a 72.1% increase in HDL cholesterol and a 24.9% decrease in LDL cholesterol at 12 months (P < 0.001). The torcetrapib arm showed an increased risk of cardiovascular events (HR, 1.25; 95% CI, 1.09–1.44; P = 0.001) and an increased risk of death from any cause (HR, 1.58; 95% CI, 1.14–2.19; P = 0.006). The study was terminated early because of these findings.	The addition of torcetrapib resulted in a substantial increase in HDL cholesterol and a marked decrease in LDL cholesterol, while also increasing the risk of mortality and morbidity through an unknown mechanism.
HER2-negative metastatic breast cancer	Progression-free survival (OS) (PFS)	Overall survival	Bevacizumab (E2100, AVADO, and RIBBON-1 studies) showed improved PFS (surrogate outcome) but no benefit in OS (primary outcome).	E2100 study ¹ : 722 patients with metastatic breast cancer. Combination therapy (paclitaxel + bevacizumab) vs. monotherapy (paclitaxel). Median PFS was 11.8 months in the combination therapy group vs. 5.9 months with paclitaxel alone (HR, 0.60; 95% CI, 0.47–0.77; P < 0.001). Mean OS was similar in the combination therapy and paclitaxel-alone groups: 26.7 months and 25.2 months, respectively (HR, 0.88; P = 0.16).	Combined paclitaxel and bevacizumab therapy increased progression-free survival and the 1-year survival rate but did not improve overall survival. The combination therapy group also had higher rates of severe toxic effects (grades 3 and 4).

HbA1c: glycated hemoglobin; T2DM: type 2 diabetes mellitus; CVR: cardiovascular risk.

Source: authors' own elaboration.

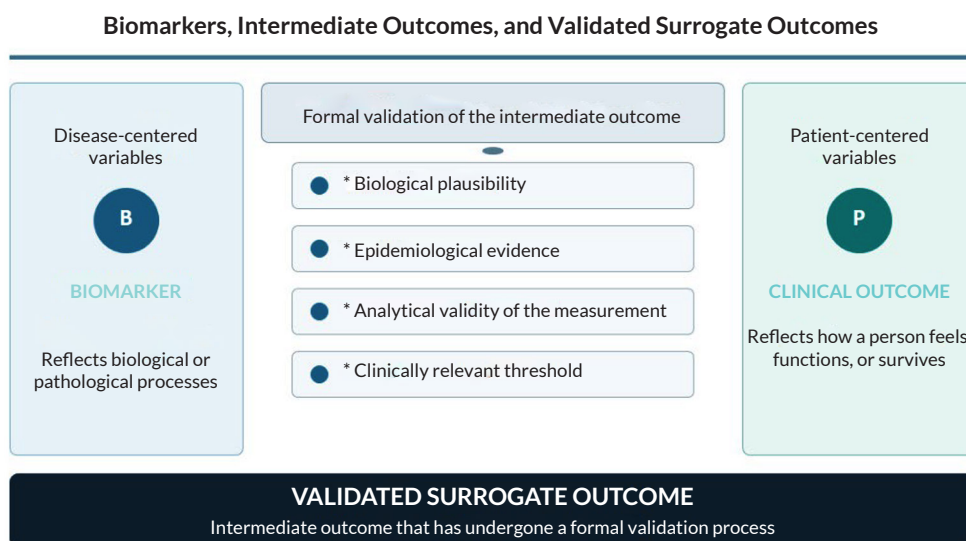


Figure 1. Biomarkers, intermediate outcomes, and validated surrogate outcomes⁶.

for a specific disease, type of intervention, and clinical outcome. Therefore, evidence obtained in one context cannot automatically be extrapolated to other clinical settings. The FDA notes that, in oncology, for example, progression-free survival depends on the type and stage of disease, the context of use, the magnitude of the effect, available treatments, and the balance between benefits and risks. Its ability to serve as a substitute for overall survival must be evaluated in each specific clinical setting.¹⁰

GENERAL CONSIDERATIONS ON VALIDATION

Various methods have been proposed to determine whether a surrogate outcome can predict a patient-centered outcome. Validation requires integrating biological plausibility, epidemiological evidence, measurement validity and reliability, and statistical evidence of its ability to predict the clinical effect of an intervention. The Prentice criteria initially established that a valid surrogate must be associated with the clinical outcome and fully capture the effect of treatment on that outcome. Because this condition is difficult to demonstrate, meta-analytic models developed by Buyse et al. assess the association between the two outcomes at the individual level and between their respective treatment effects at the trial level.

From a causal perspective, statistical methods can be grouped into causal-effects approaches, which analyze the relationship among the intervention, surrogate outcome, and clinical outcome, and causal-association approaches, which assess whether the treatment effect on the surrogate predicts the clinical effect. Among the latter, meta-analytic models examine associations at

both the individual and trial levels. Individual participant data meta-analysis is the most comprehensive approach, although its substantial data and resource requirements have also led to the use of aggregate data. Another measure is the surrogate threshold effect, which estimates the minimum change in the surrogate needed to predict a clinical effect; however, it depends on the number, quality, and heterogeneity of the available trials.¹¹

In practice, Ciani et al. propose organizing this evaluation into three "hierarchical" stages: establishing the level of available evidence, assessing the strength and consistency of these associations, and estimating the expected clinical effect based on the observed effect on the surrogate.

Because no universally accepted methods or thresholds exist, validation should be based on a progressive assessment of the totality of the evidence and should be limited to the clinical context studied.⁹

GOOD PRACTICES IN THE USE OF SURROGATE OUTCOMES

Good practices in the use of surrogate outcomes rely on several interrelated elements that together provide an integrated framework for methodological quality:

- Rigorous validation ensures that the intermediate marker has a strong relationship with the clinical outcome.¹¹
- Transparency in reporting, including adherence to CONSORT-Surrogate, makes it possible to explicitly identify the surrogate nature of the outcome, report the extent to which it has been validated, and discuss its limitations.¹²

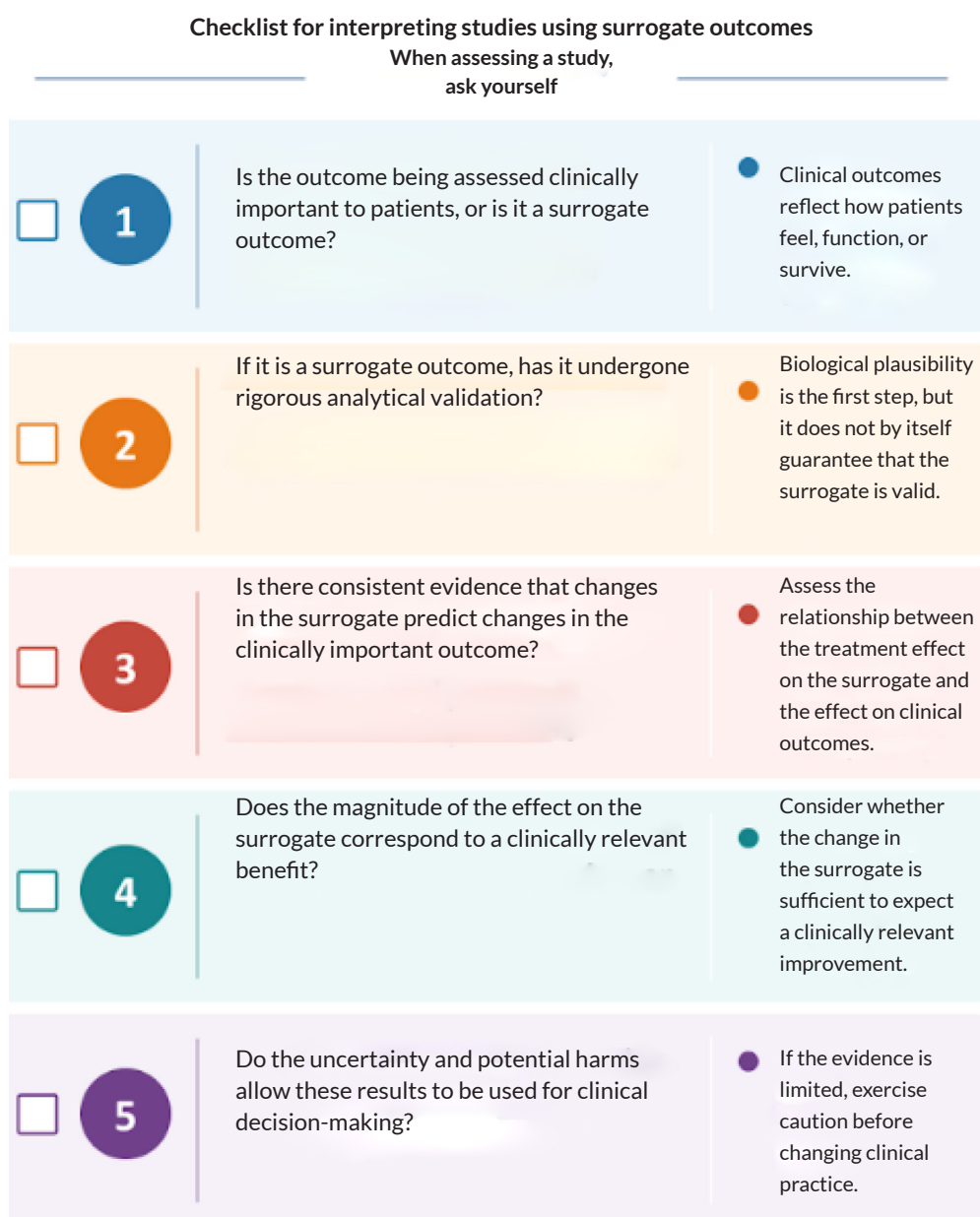


Figure 2. Checklist for interpreting studies using surrogate outcomes.¹¹

CONCLUSION

Surrogate outcomes are a useful tool in clinical research and can accelerate the development of new knowledge. However, their use without rigorous validation carries the risk of leading to the approval of interventions that are insufficiently effective, costly, or associated with a higher incidence of adverse events, without translating into meaningful improvements in patient survival or quality of life. A more critical and transparent approach to the use of surrogate outcomes will help improve the quality of evidence and healthcare decision-making.

Author contributions: Conceptualization, Project administration, Supervision, Writing – review and editing: MAB, MFGR. Funding acquisition: MFGR. Investigation, Visualization, Writing – original draft: MAB, SBS, HAPA. Methodology, Validation: MAB, SBS, HAPA, MFGR.

Conflicts of interest: The authors declare no conflicts of interest related to the content of this work.

Funding: The authors declare that this study received no funding from external sources.

REFERENCES

1. Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med*. 2007;357(26):2666-2676. DOI: <https://doi.org/10.1056/NEJMoa072113>
2. Barter PJ, Caulfield M, Eriksson M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med*. 2007;357(21):2109-2122. DOI: <https://doi.org/10.1056/NEJMoa0706628>
3. U.S. Food and Drug Administration. FDA Facts: Biomarkers and Surrogate endpoints [Internet]. Silver Spring (MD): Food and Drug Administration; 2019 [citado 2025 nov 23]. Disponible en: <https://www.fda.gov/about-fda/innovation-fda/fda-facts-biomarkers-and-surrogate-endpoints>
4. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration; Bethesda (MD): National Institutes of Health; 2016- [citado 2025 nov 23]. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK326791/>
5. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95. DOI: <https://doi.org/10.1067/mcp.2001.113989>
6. Lassere MN, Johnson KR, Boers M, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol*. 2007;34(3):607-615.
7. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ*. 2011;343:d7995. DOI: <https://doi.org/10.1136/bmj.d7995>
8. Schuster DP. The opportunities and challenges of developing imaging biomarkers to study lung function and disease. *Am J Respir Crit Care Med*. 2007;176(3):224-230. DOI: <https://doi.org/10.1164/rccm.200703-462PP>
9. Ciani O, Buyse M, Drummond M, et al. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. *ValueHealth*. 2017;20(3):487-495. DOI: <https://doi.org/10.1016/j.jval.2016.10.01110>
10. U.S. Food and Drug Administration. Guidance Documents [Internet]. Silver Spring (MD): Food and Drug Administration; [citado 2025 nov 23]. Disponible en: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents>
11. Christensen R, Ciani O, Manyara AM, et al. Surrogate endpoints: a key concept in clinical epidemiology. *J Clin Epidemiol*. 2024;167:111242. DOI: <https://doi.org/10.1016/j.jclinepi.2023.111242>
12. Manyara AM, Davies P, Stewart D, et al. Reporting of surrogate endpoints in randomised controlled trial reports (CONSORT-Surrogate): extension checklist with explanation and elaboration. *BMJ*. 2024;386:e078524. DOI: <https://doi.org/10.1136/bmj-2023-078524>
13. Churrua K, Pomare C, Ellis LA, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect*. 2021;24(4):1015-1024. DOI: <https://doi.org/10.1111/hex.13254>
14. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials*. 2017;18(Suppl 3):280. DOI: <https://doi.org/10.1186/s13063-017-1978-4>
15. Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559. DOI: <https://doi.org/10.1056/NEJMoa0802743>