

Multidimensional Validation of Clinical Evidence: A Four-Domain Audit Framework

Francesco Masedu⁽¹⁾, Monica Mazza⁽¹⁾, Margherita Attanasio⁽¹⁾, Ilenia Le Donne⁽¹⁾, Nicole Covone⁽¹⁾, Marco Valenti⁽¹⁾

 Francesco Masedu 0000-0003-0290-5324

(1) Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, Italy

CORRESPONDING AUTHOR: Francesco Masedu, Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy. E-Mail: francesco.masedu@univaq.it

INTRODUCTION

The evaluation of clinical evidence continues to rely predominantly on measures of statistical significance and relative effect size. Although indispensable, these metrics alone may not adequately characterize evidence quality, particularly with respect to follow-up integrity, clinical relevance, predictive performance, and vulnerability to bias. As regulatory and clinical decision-making increasingly incorporates evidence from both randomized and real-world studies, complementary approaches for assessing the robustness of findings are needed.

OBJECTIVES

To present a four-domain audit framework for the multidimensional validation of clinical evidence and to demonstrate its application in a proof-of-concept survival dataset.

METHODS

The proposed framework comprises four sequential domains of evaluation: (1) Follow-up Integrity, assessed through inverse Kaplan–Meier analysis to identify informative censoring and differential attrition; (2) Clinical Relevance, evaluated using Restricted Mean Survival Time (RMST) to quantify benefit in clinically interpretable time units; (3) Predictive Robustness, assessed through discrimination (Harrell's C-index), calibration analysis, and bootstrap-based optimism correction; and (4) Inferential Resilience, evaluated using E-value methodology to estimate the susceptibility of observed findings to unmeasured influences arising during study conduct and follow-up. These domains were selected to represent complementary dimensions of evidence quality: data integrity, clinical relevance, predictive validity, and inferential

robustness. The framework was applied to the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset as a proof of concept.

RESULTS

Follow-up integrity was preserved, with no evidence of differential censoring between groups (inverse Kaplan–Meier log-rank $p=0.54$). However, subsequent analyses identified important limitations. RMST analysis showed a between-group difference of 51.98 days (95% CI -341.87 to 445.84 ; $p=0.796$), indicating limited clinical relevance. Predictive performance was modest, with an optimism-corrected C-index of 0.639 and a calibration slope of 0.933. Finally, the E-value was 1.98, indicating that relatively modest unmeasured influences arising during follow-up could potentially account for the observed association. Overall, findings that appeared methodologically acceptable under conventional statistical assessment did not consistently satisfy all dimensions of the proposed validation framework.

CONCLUSIONS

The proposed audit framework provides a structured approach for integrating assessments of follow-up integrity, clinical relevance, predictive performance, and inferential resilience within a unified methodological framework. In this proof-of-concept application, preserved data integrity coexisted with limited clinical benefit, modest predictive robustness, and vulnerability to follow-up-related biases. By making these dimensions explicit and reproducible, the framework may complement conventional statistical evaluation in evidence appraisal, peer review, and regulatory assessment. Further validation across randomized and real-world datasets is warranted.