

CardioOracle: An Open, Transparent Model for Predicting Cardiovascular Trial Success

Syed Muhammad Ali

therealmoali@gmail.com

Royal Free London NHS Foundation Trust <https://orcid.org/0009-0005-7050-4401>

Mahmood Ahmad

Royal Free London NHS Foundation Trust

Short Report

Keywords: Digital health tools, ClinicalTrials.gov, Clinical trial design, Prediction models, Cardiovascular trials, Temporal validation

Posted Date: September 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11066605/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background: Cardiovascular trials require substantial investment yet frequently fail to meet their primary endpoints. Existing trial-forecasting approaches are often proprietary or difficult to interpret, limiting their value to researchers and other stakeholders. We developed CardioOracle, an open, browser-based model that estimates the probability that a cardiovascular trial will meet its primary endpoint.

Methods: Data were programmatically extracted from the Aggregate Analysis of ClinicalTrials.gov database using a deterministic pipeline. Eligible studies were randomized, interventional Phase II/III, III, or III/IV cardiovascular trials enrolling at least 50 participants and reporting publicly available results. Trials were classified as successful or unsuccessful according to whether the registered primary endpoint was met. CardioOracle combines three interpretable components: Bayesian historical borrowing from similar completed trials, conditional power informed by historical effect-size distributions, and logistic regression using prespecified trial-design characteristics. Trials completed before January 1, 2020, were used for model development and recalibration, while those completed on or after this date formed a temporal holdout set. Performance was assessed using the area under the receiver operating characteristic curve and Brier score.

Results: The final dataset comprised 784 cardiovascular trials completed between 1990 and 2024, including 133 trials in the temporal holdout set. In the overall cardiovascular cohort, CardioOracle achieved a temporal holdout area under the curve of 0.745 and a Brier score of 0.196, compared with an in-sample area under the curve of 0.787 and Brier score of 0.169. In a dedicated coronary artery disease configuration, the temporal holdout area under the curve was 0.843. Within the regression component, industry sponsorship and surrogate primary endpoints were positively associated with trial success, whereas all-cause mortality as the primary endpoint was negatively associated with success.

Conclusions: CardioOracle demonstrated moderate discrimination when applied to a temporally separated cohort of contemporary cardiovascular trials, with stronger performance in the more homogeneous coronary artery disease configuration. Its open, interpretable architecture may provide a reproducible adjunct to clinical judgment and formal statistical planning when designing and evaluating cardiovascular trials. Further external and prospective validation is required before the model can be used to inform consequential research decisions.

Introduction

As cardiovascular disease remains a leading cause of morbidity and mortality worldwide progress in treatment depends on large, complex randomised control trials. These trials require

substantial financial and participant commitment, yet many fail to meet their primary endpoint [1, 2].

For investigators, funders, and regulators, improving the efficiency of trial design remains a

major priority and challenge. Although pharmaceutical companies increasingly rely on internal

analytic tools to estimate the likelihood of trial success, these approaches are usually not publicly available and lack transparency, limiting their usefulness for academic researchers and other stakeholders.

Recent studies have explored machine learning approaches to trial outcome prediction using registry and protocol data [3], however these approaches often emphasize predictive accuracy at the expense of interpretability and are seldom made available as practical tools. Furthermore, limited studies have assessed prospective performance through temporal validation, an important step in determining real-world utility. [4, 5] To address these limitations, we developed CardioOracle, an open, browser-based tool that integrates historical data, statistical power considerations, and trial design features to predict whether a cardiovascular trial will meet its primary endpoint

Methods

Data source and trial selection

Trial metadata and results were programmatically extracted from the Aggregate Analysis of ClinicalTrials.gov (AACT) PostgreSQL database using a deterministic python pipeline. The extraction queried prespecified structured registry fields that identified interventional, randomized Phase II/II-III/IV cardiovascular trials with posted results. then assembled trial-level records for downstream analysis. No free-text scraping or generative interface was used. The extraction queried prespecified structured registry fields to assemble trial-level records for downstream analysis. The extraction and outcome-labeling pipeline is reproducible from a clean environment using the public codebase, requiring only authorized access to the AACT database.

The AACT database is a relational database designed specifically to support large-scale analyses of registered clinical trials [6]. Eligible studies were interventional, randomised Phase II/III, Phase III or Phase III/IV cardiovascular trials with enrolment of at least 50 participants and publicly posted results. Trials were then labelled as either successful or unsuccessful according to whether the registered primary endpoint was met. We used a pre-specified hierarchy that prioritised automated extraction of p-values and effect estimates, with confidence-interval heuristics and limited manual adjudication of selected landmark trials where required.

Model architecture

CardioOracle combines three complementary and interpretable components. First, a bayesian historical borrowing component estimates success probability from the outcomes of completed trials that were most similar to the target trial, based on drug class, endpoint category, comparator, population descriptors, and trial era. Secondly, a conditional power component estimates the likelihood of detecting a plausible treatment effect, informed by effect-size distributions observed in similar historical trials. Third, a logistic regression component estimates success probability using prespecified trial design features, including enrolment size, duration, blinding, sponsor type, endpoint category, and data safety monitoring board presence. The ensemble outputs a probabilistic estimate reflecting historical patterns, rather than simply a deterministic forecast. All coefficients and assumptions are transparent and inspectable by the user.

Validation

Model performance was evaluated using strict temporal validation to mimic prospective use, in line with recommended best practices for clinical prediction models [4, 5]. Temporal validation was utilized to approximate prospective performance and minimise information leakage. Trials with primary completion before 1 January 2020 were used for model development and recalibration, while trials on or after this date formed the test set. Discrimination was assessed using the area under the receiver characteristic curve (AUC), and overall accuracy using the Brier score.

Results

The final labelled dataset completed 784 cardiovascular trials between 1990 and 2024, of which 133 formed the temporal holdout set. In the full cardiovascular cohort, CardioOracle achieved an AUC of 0.745 with a Brier score of 0.196 on the temporal test set, and when compared to an in-sample AUC of 0.787 (Brier score 0.169). Performance was higher in a dedicated coronary artery disease configuration, with a temporal holdout AUC of 0.843, which was consistent with a greater homogeneity in populations and endpoints.

Concerning the regression component, industry sponsorship and use of surrogate endpoints were positively associated with trial success, whereas all-cause mortality as the primary endpoint was negatively associated. To illustrate prospective utility, CardioOracle correctly anticipated the direction of outcomes in several post-2020 trials. In one example, it assigned a high probability of success to DELIVER, a trial of dapagliflozin in heart failure with preserved ejection fraction that subsequently met its primary endpoint [7], and a lower probability to EMPACT-MI, a trial of empagliflozin following acute myocardial infarction that did not meet its primary composite endpoint [8]

Discussion

Our registry-based analysis is an open and transparent ensemble model that demonstrates moderate-to-good discrimination when applied prospectively to contemporary cardiovascular trials. By relying exclusively on publicly available data and interpretable components, CardioOracle contrasts with proprietary forecasting tools and enables reproducibility, scrutiny, and contextual interpretation of predictions. This model is intended to support, rather than replace, clinical judgement and formal statistical planning. Several limitations should be noted. Registry data may be incomplete or inconsistently reported, and the conditional power components rely solely by achievement of the primary endpoint, which may not fully capture the clinically meaningful secondary outcomes or subgroup effects. Nevertheless, the observed temporal performance suggests that historically grounded, transparent models can offer useful signals for trial design and interpretation in cardiovascular research.

Conclusion

CardioOracle is an open, browser-based tool that utilizes cardiovascular trial data to estimate the probability of future trial success. Its transparent architecture and temporal validation illustrate how digital health approaches can support future cardiovascular research planning.

Declarations

Data and code availability

Source code and labelled trial data are publicly available at <https://github.com/mahmood726-cyber/cardiooracle>

The full data extraction and outcome labeling pipeline, including registry fields queried and decision rules applied, is publicly available in the CardioOracle GitHub repository.

Acknowledgements

The authors acknowledge the use of artificial intelligence–based language assistance during manuscript preparation to improve clarity and readability. Artificial intelligence–based tools were used to assist with drafting and refining the graphical abstract; the authors reviewed and approved the final version. All scientific content, analyses, and interpretations were conceived, verified, and approved by the authors, who take full responsibility for the integrity of the work.

References

1. Wouters OJ, McKee M, Luyten J (2020) Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* 323:844–853
2. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nat Biotechnol* 32:40–51

3. Lo A, Siah KW, Wong CH (2019) Machine learning with statistical imputation for predicting drug approval. *Drug Discov Today* 24:1771–1779
4. Collins GS, Reitsma JB, Altman DG, Moons KGM (2024) Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 384:e074819
5. de Hond AAH, Leeuwenberg AM, Hooft L et al (2023) Perspectives on validation of clinical predictive algorithms. *npj Digit Med* 6:86
6. Tasneem A, Aberle L, Ananth H et al (2012) The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent enhancements. *PLoS ONE* 7:e33677
7. Solomon SD, McMurray JJV, Claggett B et al (2022) Dapagliflozin in heart failure with preserved ejection fraction. *N Engl J Med* 387:1089–1098
8. Udell JA et al (2024) Empagliflozin after acute myocardial infarction. *N Engl J Med*