Published in partnership with Seoul National University Bundang Hospital



https://doi.org/10.1038/s41746-025-01705-z

# Advancing efficacy prediction for electronic health records based emulated trials in repurposing heart failure therapies

Check for updates

Nansu Zong<sup>1,8</sup>  $\boxtimes$ , Shaika Chowdhury<sup>1,8</sup>, Shibo Zhou<sup>1,8</sup>, Sivaraman Rajaganapathy<sup>1</sup>, Yue Yu<sup>2</sup>, Liewei Wang<sup>3</sup>, Qiying Dai<sup>4</sup>, Pengyang Li<sup>5</sup>, Xiaoke Liu<sup>6</sup>, Suzette J. Bielinski<sup>2</sup>, Jun Chen<sup>2</sup>, Yongbin Chen  $\mathbb{O}^7$  & James R. Cerhan  $\mathbb{O}^2$ 

The complexities inherent in EHR data create discrepancies between real-world evidence and RCTs, posing substantial challenges in determining whether a treatment is likely to have a beneficial impact compared to the standard of care in RCTs. The objective of this study is to enhance the prediction of efficacy direction for repurposed drugs tested in RCTs for heart failure (HF). To achieve this, we propose an efficacy direction prediction framework that integrates drug-target predictions with EHR-based Emulation Trials (ET) to derive surrogate endpoints for prediction using HF prognostic markers. Our validation of the proposed novel drug-target prediction model against the BETA benchmark demonstrates superior performance, surpassing existing baseline algorithms. Furthermore, an evaluation of our framework in identifying 17 repurposed drugs – derived from 266 phase 3 HF RCTs – using data from 59,000 patients at the Mayo Clinic highlights its remarkable predictive accuracy.

Heart failure (HF) is a critical medical condition with high rates of hospitalization and mortality<sup>1</sup>, where the heart fails to adequately pump blood to meet the body's needs<sup>2</sup>. There is an urgent need for improved therapeutic approaches to effectively manage HF symptoms, enhance the quality of life, and increase survival rates. Drug repurposing is an innovative approach to drug discovery that involves finding new therapeutic uses for existing drugs<sup>3</sup>. Since the safety profiles of the drugs (e.g., pharmacokinetics/pharmacodynamics) for repurposing are already established through rigorous tests in the early phases<sup>4,5</sup>, predicting the potential efficacy direction<sup>6,7</sup> (i.e., beneficial or not compared to the current standard of care) of the repurposed drug candidate is critical to streamline the process and efficiently allocate resources for the large-scale trials like phase 3 clinical trials, which can be challenging (e.g., 20% failure rate for HF<sup>8</sup>) and costly due to multiple factors<sup>9</sup>.

A growing trend involves leveraging real-world data, such as Electronic Health Records (EHR), to simulate randomized trials (RCTs) for assessing treatment outcomes, referred to as emulated trials (ET)<sup>6,10–18</sup>. Contrary to the

controlled settings of RCTs, ETs utilize observational data from everyday clinical practices. This data encompasses the complexities of patient histories stored in databases, the absence of certain biomarkers or specific diagnostic images, and incomplete records of symptoms<sup>17</sup>. Although ETs are theoretically expected to adhere to the same eligibility criteria as RCTs, within the "target trial framework,"<sup>12</sup> to select appropriate patient groups and extract relevant data from EHR for efficacy evaluation, the discrepancies<sup>16,17</sup> between RCT standards and EHR data present a challenge in executing an ET for estimating treatment effect<sup>19–25</sup>. Although state-of-the-art ET methods<sup>26,27</sup> have incorporated AI to balance confounders and streamline patient selection criteria, their primary focus has not been on enhancing the accuracy of estimation. Furthermore, their performance in efficacy estimation—both in terms of direction and level—remains untested.

This study introduces a novel framework for predicting one of the key metrics in prioritizing repurposed drugs—efficacy direction<sup>67</sup>—which assesses whether initiating the repurposed drug is likely to be beneficial

<sup>1</sup>Department of Artificial Intelligence and Informatics Research, Mayo Clinic, Rochester, MN, USA. <sup>2</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. <sup>3</sup>Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. <sup>4</sup>Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA. <sup>5</sup>Division of Cardiology, Pauley Heart Center, Virginia Commonwealth University, Richmond, VA, USA. <sup>6</sup>Division of Community Cardiology, Department of Cardiovascular Medicine, La Crosse, Wisconsin, WI, USA. <sup>7</sup>Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, USA. <sup>8</sup>These authors contributed equally: Nansu Zong, Shaika Chowdhury, Shibo Zhou. <sup>Senanal:</sup> Zong.Nansu@mayo.edu

compared to the standard of care. Specifically, the framework integrates drug-target prediction and EHR-based ETs to derive a surrogate measure for assessing drug efficacy direction based on HF prognostic markers. For drug-target prediction, we propose a novel model that leverages drug chemical compositions, protein sequences, and heterogeneous biomedical knowledge bases, evaluated against the BETA<sup>28</sup> benchmark. For ET, we conducted virtual clinical trials using a general emulation protocol to estimate the probability of abnormal lab results for HF prognostic markers. The final efficacy direction prediction will be generated based on the integration of predicted novel drug targets with an odds-ratio-based normalization derived from a logistic regression model for abnormal prognostic biomarker risk estimation based on ET. To evaluate the proposed framework, we manually identified 17 repurposed HF drugs from 266 phase 3 RCTs-7 with a beneficial effect and 10 without. Our analysis encompassed 11 HFrelated prognostic markers: B-type Natriuretic Peptide (BNP), N-terminal pro-BNP (NT-proBNP), Troponin T (TnT), Troponin I (TnI), Renin, Aldosterone, Estimated Glomerular Filtration Rate (eGFR), C-Reactive Protein (CRP), Blood Urea Nitrogen (BUN), Creatinine, and Hemoglobin. Additionally, we considered two patient cohorts-HF with preserved ejection fraction (HFpEF) and HF with reduced ejection fraction (HFrEF)as well as HF medications (ACE inhibitors, Beta-Blockers, Angiotensin II Receptor Blockers, and Loop Diuretics) and biological factors (e.g., sex). The proposed framework significantly improved predictive accuracy compared to traditional EHR-based ET methods. Notably, While EHR-based ET can estimate both the magnitude and direction of treatment effects using odds ratios derived from the logistic regression model, our focus is solely on predicting efficacy direction.

#### Results

#### Proposed framework—integration EHR-based emulation with drug-target prediction

Our proposed an efficacy direction prediction framework that integrates drug-target prediction with EHR-based ET. The process comprises four key steps (refer to Fig. 1a for the study design): 1) identifying genes associated with the disease (i.e., HF), 2) predicting drug targets based on HF-related genes, 3) extracting the study cohort from EHR, and 4) prediction of efficacy direction based on ET and drug-target prediction.

In the initial phases, we developed an innovative algorithm for predicting targets. This algorithm is designed to use information from the chemical structure of drugs, the sequence of proteins, and biomedical relationships, with the aim of reconstructing the original knowledge graph. This reconstruction facilitates the prediction of new potential associations between drugs and their targets. In the subsequent stages, we employed a trial emulation framework<sup>29-32</sup> to predict the likelihood of achieving a normal prognostic marker during follow-up when receiving a repurposed drug, using a logistic regression model. We evaluated 11 HF prognostic markers (i.e., NT-proBNP, BNP, TnT, TnI, Creatinine, BUN, eGFR, Hemoglobin, Renin, Aldosterone, CRP) while adjusting for a wide range of variables from the Mayo Clinic (MC) database, including three biological factors-age, sex, and ethnicity-as well as various HF medications such as ACEIs, BBs, ARBs, and LDs. Additionally, we examined HF subtypes, including HFrEF and HFpEF, and incorporated BASELINE cohorts, which are specific cohorts required by the original RCTs. The MC patient database includes a total of 59,102 HF patients. For statistical details, please refer to Table 1. The flow of data generation for this study is detailed in Fig. 2. To estimate the probability of a drug's potential benefit in clinical trials (i.e., the direction of efficacy) compared to the standard of care, we normalized the odds ratio, derived from the logistic regression coefficient, and integrated it with predicted targets. For a comprehensive explanation of these methods, please see "Section Methods".

#### Evaluation of the drug-target prediction model using the BETA benchmark and prediction of 17 RCT drugs

In this section, we carry out two evaluations for our proposed drug-target prediction model (depicted in Fig. 1b): (1) an evaluation based on seven tests

and 344 tasks from the BETA benchmark, and (2) an evaluation focused on predicting 17 drugs used in HF RCTs.

For the BETA benchmark evaluation, seven tests were performed to test the model with different tasks, including general screening, target and drug screening by category, identifying specific drugs and targets, and drug repurposing for diseases. To compare our model's performance, we used six state-of-the-art predictive models (DTINet<sup>33</sup>, bioLNE<sup>34</sup>, NeoDTI<sup>35</sup>, DeepPurpose<sup>36</sup>, DeepDTA<sup>37</sup>, GraphDTA<sup>38</sup>) as baseline methods. The proposed method performed exceptionally well with an average general score of  $96.1\% \pm 3.3\%$  for Test 0,  $86.5\% \pm 17.0\%$  for Test 1,  $92.6\% \pm 8.0\%$  for Test 2,  $45.4\% \pm 26.2\%$  for Test 4,  $83.5\% \pm 9.5\%$  for Test 5,  $67.1\% \pm 14.8\%$  for Test 6 (see Fig. 3a). Our proposed method was outstanding, with 266 out of 404 tasks (65.8%) outperforming other methods across all seven tests (Supplementary Fig. 4). It's important to note that for test 4, we counted the average precision for the top 10, 20, 50, and 100, which made 80 tasks counted rather than 20 tasks counted in the BETA. For each test, our proposed method achieved excellent results with 10/10 (100.0%), 61/90 (67.8%), 66/ 72(91.7%), 62/72(86.1%), 42/80 (52.5%), 17/40 (42.5%), 8/40 (20.0%). Please refer to the Supplementary Figs. 5-16 for more detailed results in each BETA tasks.

For the prediction of 17 RCT drugs for phase 3 HF clinical trials (7 with a beneficial effect and 10 without in the original RCTs), the predictions of 25 HF-related genes are displayed in Fig. 3b. Dexamethasone, thiamine B1, ranolazine and are the top drugs predicted to have the most HF-related genes, with 24/25, 20/25, and 15/25 genes, respectively. On the other hand, Simvastatin, Warfarin, and Rosuvastatin have the least number of HFrelated genes, with only 2/25, 2/25, 2/25, and 1/25 genes, respectively. It is worth noting that all these drugs are used for treating or managing medical conditions related to the cardiovascular system. Specifically, Rosuvastatin and Simvastatin work by reducing cholesterol production and increasing clearance of LDL-cholesterol from the bloodstream by inhibiting HMG-CoA reductase, an enzyme involved in cholesterol synthesis. Regarding genes, NOS3, ADRB1, and VEGFA have the most HF-repurposed drugs, with 14/24, 12/24, and 11/24 drugs, respectively. Meanwhile, PTH, RAC 1, and GRK2 are the genes with the least number of repurposed drugs, with only 1/24 drugs for each.

# EHR-based emulation and improvement by the integration with drug-target prediction

To evaluate the efficacy of 17 repurposed drugs, we conducted a logistic regression (Refer to Section Methods for details) across 4 cohort combinations (BASELINE, ALL, HFpEF, and HFrEF) and 11 prognostic markers (NT-proBNP, BNP, TnT, TnI, Creatinine, BUN, eGFR, Hemoglobin, Renin, Aldosterone, CRP). The odds ratio was used to determine the direction of efficacy when it was significantly different from 1 (e.g., p < 0.05) in the logistic regression analysis of biomarker changes and drug intake within the EHR-based emulation. Similarly, a normalized odds ratio was combined with predicted novel targets to generate a probability score in our proposed method, assessing the likelihood of a drug being beneficial. Please note that, unlike our proposed method, which integrates the transformed odds ratio with drug-target predictions for efficacy direction estimation, the ET-based approach considers only the direction inferred from the odds ratio without incorporating its magnitude. Figure 4 highlights one of the overall best-performing combinations, specifically the combination of the BNP marker and BASELINE cohort, to represent the results of both the EHRbased emulation and the emulation integrated with drug-target prediction. The odds ratios and corresponding confidence intervals are reported in Supplementary Tables 8 and 9. Our findings show that relying solely on the EHR-based emulation does not yield satisfactory prediction for efficacy direction. While, in some cases (e.g., BASELINE and BNP), drugs with no beneficial effect (e.g., Aspirin) and those with beneficial effects (e.g., Prednisone) were accurately predicted. The majority-especially those with beneficial effects-were misclassified, resulting in high rates of false positives and negatives. However, integrating drug-target prediction significantly improved prediction accuracy at both the individual drug and overall levels.



For instance, in the BASELINE cohort and eGFR, the performance improved from AUCROC = 65.71% and PRAUC = 77.14% to AUCROC = 85.71% and PRAUC = 85.56%. For the complete set of cohort and prognostic marker combinations, refer to Supplementary Figs. 17-20.

A direct comparison between EHR-based emulation and emulation integrated with drug-target prediction is presented in Fig. 5. Panel (a)

highlights the best-performing cohorts for each prognostic marker, and a paired one-tailed t-test indicates that the emulation integrated with the drug-target prediction (average AUCROC =  $83.32 \pm 4.53\%$ ) significantly outperforms the EHR-based emulation (average AUCROC =  $55.51 \pm 10.18\%$ ) with a p-value of 5.045e-07. Panel (b) shows the pairwise comparison based on prognostic markers and HF cohorts,

**Fig. 1** | **An overview of the proposed framework. a** illustrates the general pipeline, starting with a search for relevant clinical trial drugs on ClinicalTrials.gov and identifying genes using the DisGeNET<sup>48</sup> database. With a list of repurposed trial drugs and associated genes, the proposed drug-target prediction model predicts the probabilities of drug-gene associations. We evaluate the efficacy direction of repurposed drugs—beneficial (positive impact) or not—by integrating drug-target prediction with an EHR-based emulation protocol that compares exposed and unexposed groups using HF-related clinical tests. In this framework, the odds ratio, derived from the logistic regression coefficient, serves as a predictive component and is combined with predicted targets through a normalization transformation. Importantly, while the odds ratio is used to estimate efficacy direction in the ET-based comparison method, our proposed framework normalizes the odds ratio and integrates it with predicted drug targets to generate a probability-like score, representing the drug's potential benefit in clinical trials (i.e., efficacy direction) compared

#### Table 1 | Statistics of the HF patients from the Mayo Clinic EHR

Variables	# patient	Prognostic Markers	# patient
Age ± SD	54.0 ± 31.5	NT-proBNP	5845
Ethnicity		Creatinine	58,350
Non-Hispanic White	52,329	BUN	58,315
Others	6674	Hemoglobin	58,160
Baseline HF drugs		Aldosterone	2941
ACEI	26,719	BNP	36,733
BB	46,081	CRP	35,360
ARB	18,276	eGFR	53,784
LD	53,565	Renin	2674
HF subtypes		Tnl	5989
HFpEF	24,427	TnT	37,603
HFrEF	21,134	Repurposed drugs	
Confounders		Albuterol	28,652
Arrhythmias (including baseline variations, Atrial fibrillation, Flutter)	11,566	Amiodarone	15,805
Cardiac Arrest (including history of)	726	Aspirin	41,504
Myocardial infarction (past, prior)	4492	Cholecalciferol	12,502
Hypertension (including long- standing)	19,955	Dexamethasone	26,010
Pulmonary hypertension	1680	Macitentan	200
Ischemic heart disease	10,298	Metformin	5142
Hypertensive heart and Renal disease	1597	Methotrexate	642
Medications affecting heart health	456	Mirabegron	400
Congestive heart failure, unspecified	12,074	Prednisone	15,272
Renal Failure (including end- stage, acute, and chronic)	7776	Ranolazine	834
Diabetes Mellitus (type 1 and type 2)	9083	Rosuvastatin	7611
		Sertraline	6060
		Sildenafil	771
		Simvastatin	9855
		Thiamine	5370
# HF in total	59,102	Warfarin	20,501

to the standard of care. **b** provides details about the developed drug-target prediction model, which leverages drug chemical structure, protein sequences, and a knowledge graph to generate embedding layers for each biomedical entity. The biomedical information for the same entity is concatenated to enhance prediction accuracy. For instance, drug-related information, highlighted in red, is incorporated from embeddings of drug-associated networks, such as drug-target and drug-disease bipartite networks. The cost function is optimized by minimizing the weights between reconstructed sub-knowledge graphs and the original graphs. Predictions are derived from the reconstructed drug-target graph. **c** outlines the EHR-based ET protocol developed for the study, consisting of seven core components: 1) Treatment Strategy, 2) Treatment Assignment, 3) Eligibility Criteria, 4) Follow-up, 5) Confounders, 6) Outcome, and 7) Statistical Analysis. A graphical representation of each component is provided.

where significant improvements (e.g., *p*-value < 0.01) are observed through a paired one-tailed t-test. For similar combinations based on PRAUC, refer to Supplementary Fig. 21.

In our model, the statistical odds ratio is integrated with the number of predicted targets to form a prediction score. We also evaluated a more aggressive alternative (referred to as ignoreP), which uses odds ratios without considering statistical significance. This may result in a higher or lower prediction score compared to our method, depending on the direction of the nonsignificant effects. Comparative analyses show that our proposed method achieved a better predictive performance than ignoreP with regard to AUCROC and PRAUC (see Supplementary Fig. 25).

#### Improvement by considering interaction with HF drugs and sex

For each prognostic marker, we examined interactions between HF medications and sex, emphasizing increases and decreases in performance (e.g., AUCROC) as shown in Fig. 6a. CRP, Creatinine, TnI, and BUN are the top four markers with the highest number of improved predictions when accounting for interactions with HF medications and sex. Out of the 136 improved interactions for each prognostic marker, 100% (28) of BUN interactions significantly improved when integrated with the drug-target prediction (P = 2.2e-16), 100% (28) of CREATININE interactions showed significant improvement (P = 2.2e-16), and 75% (21) of CRP interactions showed significant improvement (P = 1.185e-05). We also observed that the top four markers showed more stable improvements compared to other biomarkers, as indicated by Levene's test (F = 40.4, P = 3.0 e-9), suggesting significantly lower variance in improvement. PRAUC-based improvements are presented in Supplementary Fig. 22. Figure 6b displays detailed drugspecific improvements based on the four cohorts, demonstrating predictions across all four cohorts. Additional results for other prognostic markers can be found in Supplementary Fig. 23.

#### Discussion

To bridge the gap between RCT outcomes and those inferred from EHRs, this study introduces a novel strategy for predicting one of the key metrics in prioritizing repurposed drugs for clinical trials—efficacy direction<sup>6,7</sup> (i.e., whether the drug is beneficial compared to the standard of care)—through the integration of biological and clinical insights. Our research addresses the previously unexplored accuracy of using EHRs for drug discovery in the adaptation of ETs<sup>26,27</sup> with regard to the estimation of the direction of treatment effect. By evaluating 17 drugs repurposed for phase 3 HF RCTs, we highlighted the inherent limitations of trial simulations for efficacy direction prediction based solely on EHR data. We advocate for a more sophisticated approach based on the integration of simulations of RCTs and drug-target prediction for repurposing HF drugs (Fig. 2).

In the following sections, we will discuss several aspects and issues related to this study.

The gap between real-world evidence generation in clinical settings (e.g., EHR) and the controlled environment of RCTs means that EHR-based



Fig. 2 | The flow of the data generation for the study. A detailed data processing flow is presented to show the corresponding data size and the number of patients resulting from each processing step.

simulations cannot replicate RCT outcomes exactly<sup>6,18,39</sup>. Unlike existing studies<sup>26,27,30</sup> that aimed to improve data generation through AI methods like the propensity score (PS), our study sought to refine emulation through two primary strategies: 1) incorporating biological insights via drug-target prediction, and 2) stratification based on variable interactions (e.g., HF medications and sex). Our approach demonstrated encouraging outcomes in predicting efficacy direction, surpassing baseline ET methodologies. While our experiment reported optimal results in "Improvement by considering Interaction with HF drugs and Sex", selecting the ideal combination of interactions is not straightforward. Therefore, we believe that using KG in high-throughput drug screening could present a more viable solution. Furthermore, utilizing a more precisely tailored stratified cohort could be particularly advantageous for specific tasks in which the confounders and interacting variables within RCTs are well-defined. Developing a more sophisticated method to comprehend these variables is crucial for the process of data generation from EHR.

While our evaluation focuses on efficacy direction rather than both direction and magnitude, our ET implementation adheres to the same causal analysis framework to estimate the likelihood of achieving normal prognostic marker levels during follow-up with a repurposed drug. Unlike conventional causal analyses focused on time-to-even-i.e., the time it takes to reach an endpoint-Our study assesses the likelihood of achieving a normal prognostic marker ("event") during follow-up in patients taking a repurposed drug. This approach aligns with our endpoint, which focuses on changes in prognostic markers rather than mortality. Specifically, in timeto-event analyses, time typically refers to when an event (e.g., death) occurs. However, in our study, the timing does not necessarily capture the exact point at which the biomarker status changes (e.g., from abnormal to normal). In clinical practice, the timing and frequency of biomarker testing are often influenced by the patient's clinical condition, physician decisions, and resource availability. Therefore, rather than using a Cox or pooled logistic regression model, which models the timing or hazard rate of an event, we opted for a logistic regression model.

Conventionally, there are two complementary approaches to analyzing causal effects in ET: Per-Protocol (PP) and Intention-to-Treat (ITT) analysis. PP analyzes only participants who complete the study and adhere to the treatment protocol, while ITT includes all randomized participants, regardless of adherence or study completion. ITT-based analyses are widely used in EHR-based emulations and are advantageous for preserving randomization and reducing bias<sup>30</sup>. However, patients and their providers may often be more interested in the PP effect, as it provides insights into the impact of continuous therapy adherence, which can guide treatment decisions<sup>40</sup>. The PP approach is frequently used in noninferiority trials, where the goal is to show that a new treatment is at least as effective as the standard of care without necessarily proving superiority<sup>41,42</sup>. This makes PP analyses a valuable tool for drug repurposing studies and better fits our application scenarios for prioritizing drugs for real clinical trials. While we prioritized PP-based emulation, we are also interested in ITT analysis. Therefore, we present a comparison of PP-based and ITT-based emulation in Supplementary Fig. 24, demonstrating a moderate level of consistency in predictions between PP and ITT using EHR-based emulation. Additionally, we observe a significant level of agreement in the predicted efficacy direction when incorporating Drug-Target Prediction into ET emulation. To address the biases inherent in naïve PP analyses<sup>43</sup>, we employed IPW to adjust for adherence using baseline covariates. However, we did not include timevarying covariates<sup>43</sup> (e.g., prognostic biomarkers) due to challenges specific to EHR data, such as irregular timing and frequency of biomarker measurements influenced by patient-specific factors and resource availability. Splitting follow-up periods into fixed intervals would generate substantial missing data, necessitating imputation and introducing potential bias. By excluding time-varying covariates, we reduced imputation bias and improved data completeness and precision but acknowledged a trade-off: sacrificing some causal validity by not adjusting for post-baseline timevarying confounders. This decision reflects a balance between achieving reliable and stable estimates and maintaining accuracy in causal inference. During our implementation of ET, we also encountered challenges in defining adherence. Unlike RCTs, where adherence can be closely monitored, EHR data relies on indirect measures like prescription records and drug orders. In our analysis, adherence was defined based on these records during follow-up: patients were classified as adherent (compliers) if drug orders or prescriptions extended to or beyond the end of follow-up, and as



Fig. 3 | Evaluation of the proposed drug-target prediction model using the BETA framework and predictions for 17 repurposed RCT drugs. a illustrates the model's overall performance across all seven tests within the BETA framework, demonstrating that the proposed method significantly outperforms all baseline methods. b showcases the prediction results for 17 repurposed RCT drugs based on the

proposed model. The AUC-ROC and PR-AUC curves are provided to demonstrate the model's predictive performance. Note: \* indicates statistically significant improvement tested with a paired one-tailed t-test, where \* represents *P*-value < 0.05, \*\* for *P*-value < 0.01, \*\*\* for *P*-value < 0.001, and \*\*\*\* for *P*-value < 0.001.

repurposed drug through simulation. Additionally, we incorporated base-

non-adherent (noncompliers) otherwise. While this measure captures ongoing treatment activity, it does not ensure that patients consistently take their prescribed medications. To address this limitation, we incorporated the frequency of medication orders as a proxy for ongoing drug use, capturing how regularly HF-related drugs are prescribed during the follow-up. Recognizing the limitations of accurately capturing adherence with EHR data, we believe the frequency of medication orders serves as a practical indicator of whether patients are likely to maintain therapy over time and their level of access to medical facilities. This variable was incorporated into our model to help mitigate the adherence and variations in medical access.

HF is a prominent health concern affected by a range of factors, including genetics, environmental influences, and various chronic and acute conditions<sup>44–47</sup>. The drugs we evaluated from phase 3 trials were specifically chosen for their potential to address the underlying causes of HF and to alleviate symptoms, targeting different HF categories such as HFpEF, HFrEF, or HF in general. These trials often employed a variety of selection criteria, taking into account the study's goals, potential risks, and confounding factors. In our research, we utilized demographic information, HF medications, HF subtypes and confounding variables to estimate the likelihood of achieving normal prognostic marker levels when taking a

line values of prognostic markers into our logistic regression model to account for patients' historical or current abnormal marker levels. This approach aimed to adjust for pre-existing differences without excluding individuals with abnormal baseline values, as such exclusions could have introduced selection bias and reduced statistical power. Patients with abnormal baseline values often represent a population with more severe conditions or unique treatment trajectories. Excluding these individuals could skew the analysis, potentially overestimating or underestimating the true effect of the repurposed drug. By including baseline abnormal values, we adopt a balanced approach that retains the full spectrum of patient representation in the dataset while appropriately adjusting for differences between exposed and unexposed groups. However, we recognize the limitations of relying solely on baseline values, failing to capture the dynamic nature of prognostic markers over time. Unfortunately, the challenges in generating reliable time-varying data points remain due to the irregular timing and frequency of biomarker measurements in EHR data. Future work should explore the application of advanced modeling techniques to extract and utilize time-varying data points while balancing trade-offs among data completeness, reduced bias, more accurate causal inference, and



Fig. 4 | Evaluation of EHR emulation with and without the integration of drugtarget prediction. a displays all analyses conducted using the BASELINE cohort across 11 prognostic outcome measures (NT-proBNP, BNP, TnT, TnI, Creatinine, BUN, eGFR, Hemoglobin, Renin, Aldosterone, CRP). Two subplots are included: (i) EHR-based emulation and (ii) EHR emulation with integrated drug-target prediction. Detailed predictions for each of the 17 RCT drugs are shown, along with the overall AUC-ROC and PR-AUC for all drugs. Additionally, **b** presents analyses based on BNP, the bestperforming prognostic outcome measure, across 4 HF patient cohorts (BASELINE, ALL,

HFpEF, HFrEF), with two subplots: (i) EHR-based emulation and (ii) EHR emulation with integrated drug-target prediction. Please note, for (i) EHR-based emulation, while the odds ratio derived from the logistic regression coefficient can be used to estimate both the magnitude and direction of treatment effects, we focus solely on estimating the direction to align with the objectives of this study. Similarly, (ii) For EHR emulation with integrated drug-target prediction, our model predicts the likelihood of a drug being beneficial compared to the standard of care.

generalizability. These refinements could enhance the precision and reliability of causal inference in the context of observational data and better emulate the conditions of randomized controlled trials.

The drug-target prediction process is essential for evaluating the potential of repurposing drugs for new therapeutic uses. In prioritizing genes that might be targeted by drugs, our proposed method utilizes GDA scores<sup>48</sup>—derived from various scientific sources including the volume of supporting literature—as criteria for gene selection, treating all genes as equally in the prediction of drug-target relationships. Yet, the relevance of specific genes should be assessed based on their role in biological processes, such as those involved in pharmacokinetics and

pharmacodynamics, to tailor treatments more closely to individual needs. For instance, previous research has indicated that genes like AGT and HIF1A are notably linked to an increased risk of HFpEF in individuals with chronic kidney disease<sup>49</sup> and those suffering from obesity and metabolic syndrome<sup>50</sup>, respectively. Although the significance is based on the prediction of the potential novel drug-target associations, it may be beneficial to evaluate these genes more comprehensively as they are discovered to play a more critical role in drug development (e.g., AGT plays a central role in the regulation of blood pressure<sup>51</sup>, and HIF1A has been implicated in the development of cardiac hypertrophy and fibrosis at a low oxygen level<sup>52</sup>).



Fig. 5 | Comparison of AUCROC between the EHR-based emulation and the ET integrated with drug-target prediction. a displays the comparison for each prognostic outcome measure, with a focus on the best-performing HF cohort, where the EHR-based emulation is highlighted in red, and the ET integrated with drug-target prediction is highlighted in blue. b shows the comparison based on (i) outcome measures and (ii) HF

cohorts, respectively. As shown in (**b**), a significant improvement was observed in the ET integrated with drug-target prediction. Note: \* indicates statistically significant improvement tested with a paired one-tailed t-test, where \* represents *P*-value < 0.05, \*\* for *P*-value < 0.01, \*\*\* for *P*-value < 0.001, and \*\*\*\* for *P*-value < 0.001.

Our study was conducted solely based on MC data, but there are more sophisticated approaches that can be explored by leveraging multiple EHR sources. There are two potential directions to consider. Firstly, utilizing integrated EHR databases, such as All of Us<sup>53</sup>, could offer a feasible solution to analyze a large patient cohort. However, this approach may introduce potential bias as the factors used to pull the cohort may vary across different participating hospitals. Secondly, sharing models, like federated learning<sup>54,55</sup>, could be employed, where separate models are built with a selected set of hospitals within an existing research network, such as PCORnet<sup>56</sup>. However, this approach may require additional effort in data processing and could face common challenges, such as the competitive nature of maintaining advantages among participating sites<sup>57,58</sup>. Furthermore, although a more advanced ad-hoc linear model could be utilized to select the combination of the cohort and potentially achieve better prediction performance, we did not include it in the results section due to concerns about overfitting and lack of generalizability.

The focus of our study was to assess whether a drug is likely to be beneficial—evaluating the direction of efficacy rather than estimating treatment effects. Treatment effects encompass both magnitude and direction. Given the diversity of clinical trial endpoints, estimating treatment effects across a large number of drugs poses challenges in feasibility and rigor. This process requires access to complete published analyses and



Fig. 6 | Results of incorporating interactions between HF drugs and gender in the proposed method (ET integrated with Drug-Target Prediction) for the top four prognostic markers that yielded the highest number of improved predictions. a Illustrates the performance of predictions based on various interaction combinations for the top four prognostic markers: (i) CRP, (ii) Creatinine, (iii) TnI, and

(iv) BUN. Red nodes indicate an improvement in AUCROC, while blue nodes represent a decrease in AUCROC performance. The size of each node corresponds to the AUCROC score. **b** Provides a detailed view of the improved predictions for 17 drugs across the four prognostic markers.

the availability of corresponding clinical endpoints in EHRs. However, certain endpoints or outcomes are often missing from routine clinical records. For example, exercise-based measurements and functional assessments (e.g., Change in 20 Watt Exercise Pulmonary Vascular

Resistance in NCT02885636) are typically recorded only in clinical trials and not in standard EHRs. Replacing them with surrogate or proxy endpoints (e.g., alternative HF prognostic markers) or excluding such studies introduces challenges in the interpretation and justification of causal effects, as well as the generalizability of the methods. In this study, we focused on a feasible objective—predicting efficacy direction. Our framework can serve as a high-throughput drug screening tool, helping to prioritize promising drug candidates for further clinical investigation. Importantly, our method does not replace clinical trials, as it is not designed to simulate clinical trials for exact treatment effect estimation. We recognize an interesting future direction- investigating whether incorporating biological insights (e.g., predicting novel drug targets or pathways) could enhance treatment effect estimation. Additionally, while our initial case study based on 17 drugs demonstrates promising predictive performance, drawing definitive conclusions about the method's effectiveness at this stage would be premature. Therefore, it is essential to evaluate whether the proposed method can achieve similarly promising results across a broader range of diseases. This assessment will refine the method's effectiveness and support its application in drug repurposing for precision medicine in clinical settings.

#### Methods

#### Repurposing drugs for HF in phase 3 RCTs

We conducted a search on ClinicalTrials.gov using the keywords "heart failure" and filtered by "completed" status and "phase 3" phase in order to acquire drugs undergone the phase 3 RCTs. Out of the total 266 clinical trials gathered, two experts, Drs. Chen and Dai, conducted reviews of each trial to assess whether it was specifically designed for HF and evaluated the repurposed efficacy. Their evaluations were based on the trial's results or related articles associated with the clinical trial number. The reviews provided by both experts were subsequently merged for comparative analysis, and only those trials exhibiting consistent determinations were ultimately included. After this review, we narrowed down the list to 80 clinical trials for 53 drugs. Subsequently, an additional filtering process was conducted to exclude drugs that were not present in the BETA benchmark. This led to the identification of a list comprising 17 drugs. Out of the drugs we examined, 7 were determined to have beneficial impacts on HF, while the other 10 showed no beneficial effects. For a comprehensive list of repurposed drugs, please refer to Supplementary Table 2.

#### Retrieve of the potentially druggable genes

Using the publicly available knowledgebase, DisGeNET<sup>48</sup>, we conducted a search for genes associated with HF. DisGeNET is a comprehensive database that contains genes and variants linked to human diseases, sourced from GWAS catalogs, scientific literature, and animal models. By entering the search term "heart failure" (UMLS CUI: C0018801) into the DisGeNET platform, we obtained a list of 1499 results. We filtered them based on the GDA<sup>48</sup> score, which is a confidence score calculated by the number and type of sources (level of curation, organisms). We identified 25 genes related to HF that were also present in the BETA benchmark. For more information on these genes, please refer to Supplementary Table 3.

#### **Drug-target prediction**

Given a heterogeneous network G(V, E), a set of vertices V (i.e., biomedical entities), and a set of edges E (i.e., known associations), where V are multiple types of vertices (e.g., drugs or targets) and E are multiple types of edges that connect the vertices (e.g., drug-target associations), our objective is to predict the potential new associations among V. Specifically, we have four types of vertices, which are drugs D, targets T, diseases Ds, and side effects S, where  $D \in V, T \in V, Ds \in V, S \in V$ . For each drug d, a chemical structure of such a drug is given as  $str_d$ . For each target t, a protein sequence of such a target is given as  $seq_i$ . When there is a linking edge existing in the G, the pair of vertices is defined as  $\{(u, v)|u \in V \land v \in V \land (d, t) \in E\}$ , while there is an unknown linking edge, the pair of vertices is defined as  $\{(u, v)|u \in V \land v \in V \land (d, t) \notin E\}$ . Therefore, the problem of drug-target prediction can be defined as, given a pair of a drug d and a target t, predicting whether the pair  $\{(d, t)|d \in V \land t \in V\}$  is an existent association (referred to as positive) or nonexistent association (referred to as negative).

We have developed a novel hybrid model for drug-target prediction with the incorporation of two encoding networks (as depicted in Fig. 1b), the D-network that learns the biological features from drug chemical structure and protein sequence, the N-network that learns the topological features from a set of bipartite knowledge graphs. Specifically, for the D-network, we get the latent vectors (i.e., embeddings) generated from the encoder  $\mathscr{R}_{drug}(str_d)$  for the drug's chemical structure. Similarly, the latent vectors for protein sequences are obtained from the encoder  $\mathcal{R}_{tareet}(seq_t)$ . In practice, we adopted the encoder layers from DeepPurpose<sup>36</sup>. For the N-network, we get the embedding of a node v from the t-th layer as  $\bar{f}^t(v) = \sum_{e(v,v') \in E} f^{t-1}(v'), f^{t-1}(v')$  embeddings of the neighborhood v'. In practice,  $\overline{f^t}(v)$  is implemented based on NeoDTI<sup>35</sup>. We further updated the embedding as  $f^{t}(\nu) \leftarrow Concat(f^{t-1}(\nu), \mathcal{R})$ , where  $\mathcal{R}$  is an embedding learned from D-network. Specifically,  $\mathcal{R} = \mathcal{R}_{drug}$  when  $\nu$  is a drug and  $\mathcal{R} = \mathcal{R}_{target}$  when  $\nu$  is a target. The output of the designed network is  $\mathbb{C}$ which corresponds to the set of bipartite graphs. The loss function is  $\mathbb{C} = \lambda_1 \mathbb{C}_1 + \lambda_2 \mathbb{C}_q \dots + \lambda_k \mathbb{C}_k$ , where  $\lambda_1 + \lambda_2 + \dots + \lambda_k = 1$ .  $\mathbb{C}_k$  the objective function for the subgraph k, which is calculated:  $\mathbb{C}_k = \sum_{e(u,v) \in E} ||s(e) - f(u)\mathbb{QN}^T f(v)||^2$ , where s(e) is the weight of an edge e and s(e) = 1 if there is a link between the nodes u and v. f(u) is the embedding of node u,  $\mathbb{Q}$  and  $\mathbb{N}$  are the edge projection function. For a detailed description of the proposed model, please refer to Supplementary Note 1. For the evaluation of drug-target prediction, please refer to Section 'Evaluation of the Drug-Target Prediction Model Using the BETA Benchmark and Prediction of 17 RCT Drugs'.

# Prediction of efficacy direction based on ET and drug-target prediction

Since our objective is to approximate RCT conditions and evaluate the efficacy direction of a drug, we focus on two key steps: 1) Estimating the odds ratio-derived from the logistic regression coefficient-to quantify the association between initiating a repurposed drug and the likelihood of achieving a normal prognostic marker during follow-up in emulation, and 2) Incorporating a normalization transformation of the odds ratio along with predicted novel targets to predict the prediction of efficacy direction. Specifically, ET addresses the question: What is the odds ratio associated with taking a repurposed drug and the probability of achieving a normal prognostic marker? To answer this, we conducted an emulation based on PP and intention-to-treat ITT analyses for each of the 17 repurposed RCT drugs, following an adaptation of the trial emulation framework<sup>29-32</sup>. The analysis was stratified based on patients who strictly adhered (for PP) or did not adhere (ITT) to the RCT protocol<sup>41,42</sup>. The EHR data of MC HF patients used in this study were obtained with appropriate approval from the MC's Institutional Review Board (IRB), under approval reference number IRB:23-009802. The need for informed consent was waived by the IRB as the study has no direct interaction with participants.

The ET follows the protocol detailed in Supplementary Note 2, which outlines key aspects such as eligibility criteria, treatment strategies, treatment assignment, follow-up periods, confounders, outcomes, and statistical analysis. Specifically, eligibility was predefined for general HF patients who had received standard-of-care treatment with HF medications prior to being or not being administered the targeted repurposed drugs. HF patients were identified using ICD-9 and ICD-10 codes (see Supplementary Table 4), and four medications recommended by the AHA/ACC/HFSA guidelines<sup>47</sup>—ACEI, BB, ARB, and LD—were considered as baseline HF medications for the unexposed cohort (see Supplementary Table 5).

Patients were divided into two groups:

- The treatment group (exposed), consisted of patients who received continuous HF medications along with targeted repurposed drugs during the follow-up period.
- The standard-of-care group (unexposed), consisted of patients who received continuous HF medications but did not receive any targeted repurposed drugs during follow-up.

Time zero was defined as the point when the eligibility criteria were met, and treatment was assigned by healthcare providers. Follow-up started at time zero, with the standard-of-care group receiving any standard HF drugs and the treatment group receiving a targeted repurposed drug. We predefined different minimum follow-up durations for each drug, based on the original RCTs for the 17 drugs studied (see Supplementary Table 6 for follow-up details).

Patients were stratified based on their HF clinical subtype, either HFrEF or HFpEF. In addition to these static subtypes, we also incorporated a dynamically stratified group called the BASELINE group, which was defined according to the eligibility criteria specified in the RCTs (see Supplementary Table 2 for details on the BASELINE cohort for each drug). Beyond age, sex, and ethnicity, we included 11 comorbidities and 4 HF medications as baseline confounders. These were defined by domain experts (Drs. Li, Dai, Chen, and Bielinski) to minimize bias in this study. To further minimize potential bias, we introduced two additional confounders: 1) Length of HF medication consumption, reflecting the duration patients received HF medication prior to each prognostic marker measurement, and 2) Frequency of medication orders reflects varying levels of adherence to treatment and access to HF-related services. Longer durations of HF medication use may result in greater improvements in biomarkers, while differences in adherence and access to HF-related resources can also affect prognostic marker outcomes. Including these two variables enables a more precise evaluation of the repurposed drugs' impact on biomarker changes.

We assessed 11 clinical tests commonly used as HF prognostic markers<sup>44,47,59,60</sup>, including BNP, NT-proBNP, TnT, TnI, Renin, Aldosterone, eGFR, CRP, BUN, Creatinine, and Hemoglobin. The clinical results were categorized as either normal or abnormal, based on standard clinical ranges (see Supplementary Table 7 for biomarker categories<sup>60</sup> and ranges). In order to mitigate bias, we incorporated baseline prognostic marker values at time zero and accounted for the elapsed time—defined as the interval from time zero to the latest biomarker measurement—into our model. Please note that, as the timing and frequency of biomarker testing often depend on the patient's clinical condition, physician decisions, and resource availability, including this elapsed time as confounders helps account for the variability in measurement schedules and clinical decision-making. For detailed cohort information, please refer to Table 1.

Given that the advanced method to compute the PS did not yield improved results in balancing variables<sup>27</sup>, we adopted the traditional logistic regression-based PS matching model to balance the confounders before the analysis. Specifically, the PS is generated based on the logistic regression in the prediction of the drugs, which is further used to generate the pair-wised exposed/unexposed groups for analysis<sup>61</sup>. Additionally, to address adherence-related biases, we employed IPW to weight patients based on their adherence levels, categorized as either completing or not completing follow-ups. Specifically, IPW was implemented using baseline covariates, without incorporating time-varying variables (e.g., prognostic markers) due to the substantial missing data generated when splitting follow-up periods into fixed time intervals. Further discussion on this limitation is provided in the Discussion section. We employed a logistic regression model, adjusted for relevant confounders, to quantify the probability of achieving a normal prognostic marker in exposed versus unexposed groups. The odds ratio, derived from the logistic regression coefficient, represents the association between initiating the repurposed drug and the likelihood of achieving a normal prognostic marker.

To further predict the efficacy direction, we incorporated a normalized transformation of the odds ratio and combined it with predicted novel targets using the following equation:  $score = Norm(\frac{1}{1+e^{-creffeett}} * \#target)$ , where #target is the number of positive targets predicted and Norm(\*) is a normalized function. In our proposed model, only statistically significant odds ratios (e.g., p < 0.05) are used to compute the prediction score, while nonsignificant ones are treated as having a neutral effect to avoid overinterpretation of uncertain emulation results. In contrast, the ignoreP approach incorporates all odds ratios regardless of statistical significance. This can impact the final prediction score even when the emulated effect is not statistically significant. It is important to distinguish between the role of emulation and our study objective: In the emulation process, the odds ratio —derived from the logistic regression coefficient—serves as an estimate of

the association between taking a repurposed drug and the likelihood of achieving a normal prognostic marker, capturing both its direction and magnitude. However, for efficacy direction prediction, we transform the odds ratio and integrate it with drug-target predictions to generate a normalized probability-like score. This score does not quantify the magnitude of treatment efficacy but instead represents the likelihood that the drug exhibits beneficial efficacy compared to the standard of care.

To calculate AUCROC and PRAUC, we compared the normalized prediction scores from the proposed model against the labeled drugs (i.e., a beneficial effect or not).

#### **Data availability**

The biomedical knowledge graph, BETA, is available at https://github.com/ bioIKEA/IKEA\_BETA\_Benchmark. In accordance with the Mayo Clinic's data-sharing policy to protect human subjects, we are unable to share patient-level EHR data directly. For any requests to access this EHR data, please contact the corresponding author, N.Z.

#### Code availability

The used code is publicly available at https://github.com/bioIKEA/BETA\_ Trial\_Prediction.

Received: 29 March 2024; Accepted: 3 May 2025; Published online: 24 May 2025

#### References

- 1. Shah, K. S. et al. Heart failure with preserved, borderline, and reduced ejection fraction. *J. Am. Coll. Cardiol.* **70**, 2476–2486 (2017).
- 2. Groenewegen, A., Rutten, F. H., Mosterd, A. & Hoes, A. W. Epidemiology of heart failure. *Eur. J. heart Fail.* **22**, 1342–1356 (2020).
- 3. Park, K. A review of computational drug repurposing. *Transl. Clin. Pharmacol.* **27**, 59–63 (2019).
- Wang, X. & Guan, Y. COVID-19 drug repurposing: a review of computational screening methods, clinical trials, and protein interaction assays. *Med. Res. Rev.* 41, 5–28 (2021).
- Talevi, A. & Bellera, C. L. Challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics. *Expert Opin. Drug Discov.* 15, 397–401 (2020).
- Franklin, J. M. et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. *Circulation* 143, 1002–1013 (2021).
- Wallach, J. D. et al. Assessing the use of observational methods and real-world data to emulate ongoing randomized controlled trials. *Clin. Trials* 20, 689–698 (2023).
- Bayes-Genis, A. et al. Biomarkers in heart failure clinical trials. A review from the Biomarkers Working Group of the Heart Failure Association of the European Society of Cardiology. *Eur. J. Heart Fail.* 24, 1767–1777 (2022).
- bookS Metaphor, S. & StatiSticS, A. Time for one-person trials. *Nature* 520, 30 (2015).
- Ling, Y., Upadhyaya, P., Chen, L., Jiang, X. & Kim, Y. Emulate Randomized Clinical Trials using Heterogeneous Treatment Effect Estimation for Personalized Treatments: Methodology Review and Benchmark. *J. Biomed. Inform.* 104256 (2022).
- Labrecque, J. A. & Swanson, S. A. Target trial emulation: teaching epidemiology and beyond. *Eur. J. Epidemiol.* 32, 473–475 (2017).
- Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
- Matthews A. A., Danaei G., Islam N., Kurth T. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ* 378, e071108 (2022).
- Kutcher, S. A., Brophy, J. M., Banack, H. R., Kaufman, J. S. & Samuel, M. Emulating a randomised controlled trial with observational data: an

introduction to the target trial framework. *Can. J. Cardiol.* **37**, 1365–1377 (2021).

- Lenain, R. et al. Clinical trial emulation by matching time-dependent propensity scores: the example of estimating impact of kidney transplantation. *Epidemiology* **32**, 220–229 (2021).
- Groenwold, R. H. Trial emulation and real-world evidence. JAMA Netw. Open 4, e213845–e213845 (2021).
- Franklin, J. M., Glynn, R. J., Suissa, S. & Schneeweiss, S. Emulation differences versus biases when calibrating RWE findings against RCTs. *Clin. Pharmacol. Ther.***107**, 735 (2020).
- Wang, S. V. et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 329, 1376–1385 (2023).
- Emmanuel, T. et al. A survey on missing data in machine learning. J. Big Data 8, 1–37 (2021).
- 20. Marlin B. M. Missing data problems in machine learning. (2008).
- Jerez, J. M. et al. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif. Intell. Med.* 50, 105–115 (2010).
- Lakshminarayan K., Harp S. A., Goldman R. P., Samad T. Imputation of missing data using machine learning techniques. In: *Proc. Second International Conference on Knowledge Discovery and Data Mining* KDD'96 (1996).
- Batista, G. E. & Monard, M. C. An analysis of four missing data treatment methods for supervised learning. *Appl. Artif. Intell.* 17, 519–533 (2003).
- 24. Richman M. B., Trafalis T. B., Adrianto I. Missing data imputation through machine learning algorithms. In: *Artificial Intelligence Methods in the Environmental Sciences* (Springer, 2009).
- Bertsimas, D., Pawlowski, C. & Zhuo, Y. D. From predictive methods to missing data imputation: an optimization approach. *J. Mach. Learn. Res.* 18, 1–39 (2018).
- Liu, R., Wei, L. & Zhang, P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* 3, 68–75 (2021).
- Zang, C. et al. High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data. *Nat. Commun.* 14, 8180 (2023).
- Zong, N. et al. BETA: a comprehensive benchmark for computational drug-target prediction. *Brief. Bioinform.* 23, 1–16 (2022).
- Honap S., Danese S., Peyrin-Biroulet L. Target trial emulation: improving the quality of observational studies in inflammatory bowel disease using the principles of randomized trials. *Inflamm. Bowel Dis.* 31, 843–849 (2024).
- Scola, G. et al. Implementation of the trial emulation approach in medical research: a Scoping review. *BMC Med. Res. Methodol.* 23, 186 (2023).
- Sheffield, K. M., Dreyer, N. A., Murray, J. F., Faries, D. E. & Klopchin, M. N. Replication of randomized clinical trial results using real-world data: paving the way for effectiveness decisions. *J. Comp. Effect. Res.* 9, 1043–1050 (2020).
- Shin, J.-I. & Grams, M. E. Trial emulation methods. Am. J. Kidney Dis. 83, 264–267 (2024).
- 33. Luo, Y. et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **8**, 573 (2017).
- Zong, N. et al. Drug-target prediction utilizing heterogeneous bio-linked network embeddings. *Brief. Bioinform.* 22, 568–580 (2021).
- Wan, F., Hong, L., Xiao, A., Jiang, T. & Zeng, J. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. *Bioinformatics* 35, 104–111 (2019).
- Huang, K. et al. Deep Purpose: a deep learning library for drug-target interaction prediction. *Bioinformatics*, **36**, 5545–5547 (2020).

- Öztürk, H., Özgür, A. & Ozkirimli, E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 34, i821–i829 (2018).
- Nguyen, T., Le, H. & Venkatesh, S. GraphDTA: prediction of drug-target binding affinity using graph convolutional networks. *Bioinformatics*, **37**, 1140–1147 (2019).
- 39. Boyne, D. J. et al. Association of a shortened duration of adjuvant chemotherapy with overall survival among individuals with stage III colon cancer. *JAMA Netw. Open* **4**, e213587–e213587 (2021).
- 40. Hernán, M. A. & Hernández-Díaz, S. Beyond the intention-to-treat in comparative effectiveness research. *Clin. Trials* **9**, 48–55 (2012).
- Tripepi, G., Chesnaye, N. C., Dekker, F. W., Zoccali, C. & Jager, K. J. Intention to treat and per protocol analysis in clinical trials. *Nephrology* 25, 513–517 (2020).
- 42. Sedgwick P. Intention to treat analysis versus per protocol analysis of trial data. *BMJ* **350**, (2015).
- Murray, E. J. & Hernán, M. A. Adherence adjustment in the Coronary Drug Project: a call for better per-protocol effect estimates in randomized trials. *Clin. Trials* **13**, 372–378 (2016).
- 44. Bozkurt, B. et al. Universal definition and classification of heart failure: a report of the heart failure society of America, heart failure association of the European society of cardiology, Japanese heart failure society and writing committee of the universal definition of heart failure. *J. Card. Fail.* **27**, 387–413 (2021).
- Dunlay, S. M. et al. Prognostic value of biomarkers in heart failure: application of novel methods in the community. *Circulation: Heart Fail.* 2, 393–400 (2009).
- Kittleson, M. M. et al. 2023 ACC Expert Consensus decision pathway on management of heart failure with preserved ejection fraction: a report of the American College of Cardiology Solution Set Oversight Committee. J. Am. Coll. Cardiol. 81, 1835–1878 (2023).
- Heidenreich, P. A. et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J. Am. Coll. Cardiol. 79, e263–e421 (2022).
- Piñero, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* gkw943 (2016).
- 49. Xu, Y., Rong, J. & Zhang, Z. The emerging role of angiotensinogen in cardiovascular diseases. *J. Cell. Physiol.* **236**, 68–78 (2021).
- Warbrick, I. & Rabkin, S. W. Hypoxia-inducible factor 1-alpha (HIF-1α) as a factor mediating the relationship between obesity and heart failure with preserved ejection fraction. *Obes. Rev.* 20, 701–712 (2019).
- 51. McMurray, J. J. et al. Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N. Engl. J. Med.* **371**, 993–1004 (2014).
- 52. Krishnan, J. et al. Activation of a HIF1 $\alpha$ -PPAR $\gamma$  axis underlies the integration of glycolytic and lipid anabolic pathways in pathologic cardiac hypertrophy. *Cell Metab.* **9**, 512–524 (2009).
- 53. Investigators AoURP. The "All of Us" research program. *N. Engl. J. Med.* **381**, 668–676 (2019).
- 54. Xu, J. et al. Federated learning for healthcare informatics. *J. Healthc. Inform. Res.* **5**, 1–19 (2021).
- 55. Yang, Q. et al. Federated learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **13**, 1–207 (2019).
- Fleurence, R. L. et al. Launching PCORnet, a national patientcentered clinical research network. J. Am. Med. Inform. Assoc. 21, 578–582 (2014).
- Cifuentes, M. et al. Electronic health record challenges, workarounds, and solutions observed in practices integrating behavioral health and primary care. *J. Am. Board Fam. Med.* 28, S63–S72 (2015).
- Reisman, M. EHRs: the challenge of making electronic data usable and interoperable. *Pharm. Ther.* 42, 572 (2017).
- Gaggin, H. K. & Januzzi, J. L. Jr. Biomarkers and diagnostics in heart failure. *Biochim. Biophys. Acta (BBA)-Mol. Basis Dis.* 1832, 2442–2450 (2013).

 Westreich, D., Lessler, J. & Funk, M. J. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J. Clin. Epidemiol.* **63**, 826–833 (2010).

### Acknowledgements

This work is supported by a grant from the National Institute of Health (NIH) NIGMS (R00GM135488).

## Author contributions

N.Z. conceptualized, supervised, and managed the whole study. NZ designed the drug-target predictive model, with S.C. and S.Z. leading the development and optimization of the model. N.Z. designed and developed the emulation framework, with S.C. processing the EHR data. N.Z. conducted experiments to test drug-target prediction and emulations. N.Z. performed data visualization, with S.R. and Y.Y. contributing to plotting. Q.D., Y.C., and S.B. reviewed the confounding list, with annotation by Y.C. and Q.D., and subsequently reviewed it by N.Z., P.L., and Y.C. Statistical analysis methods were provided by N.Z., P.L., and J.C. L.W., X.L., and J.R.C. provided insights and comments on methods and experiments. N.Z. composed the initial draft of the report and all the responses to the reviewers' comments. The final manuscript was reviewed, enhanced, and approved by all authors. N.Z. assumed collective responsibility for deciding to submit the report for publication, addressing any concerns regarding accuracy and integrity.

## **Competing interests**

The authors declare no competing interests.

# Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-01705-z.

**Correspondence** and requests for materials should be addressed to Nansu Zong.

Reprints and permissions information is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/bync-nd/4.0/.

© The Author(s) 2025