因果读书会论文解读分享：机器学习与医疗

因果机器学习与医疗：在观察性数据中的应用

报告人：赵欣

UCL数学系 帝国理工公共卫生学院 毕业生

初创公司 研究科学家

ARTICLES
https://doi.org/10.1038/s42256-020-00276-w
nature machine intelligence
Check for updates

A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data

Ruoqi Liu[1], Lai Wei[2] and Ping Zhang[1,2,3]✉

Liu et al. (2021): POM framework

Outline

- Background
  - Why repurposing drugs?
  - Conventional approaches to drug repurposing
  - Regulatory incentives
  - Study aim
- Methods
  - Dataset
  - Pre-processing
  - Model for propensity score (PS) estimation
  - Average treatment effect
  - Evaluation metrics
- Results
  - Cohort characteristics
  - Feature balancing and ATE estimates
  - Comparison of expriments
- Discussions
  - Main findings
  - Strengths and limitations
  - Future work

Background: why repurposing drugs?

- Shortened drug development
  - No safety testing required on animals or healthy people, can jump straight into clinical trials to understand treatment effect
- Efficiency and feasibility
  - Particularly for rare diseases

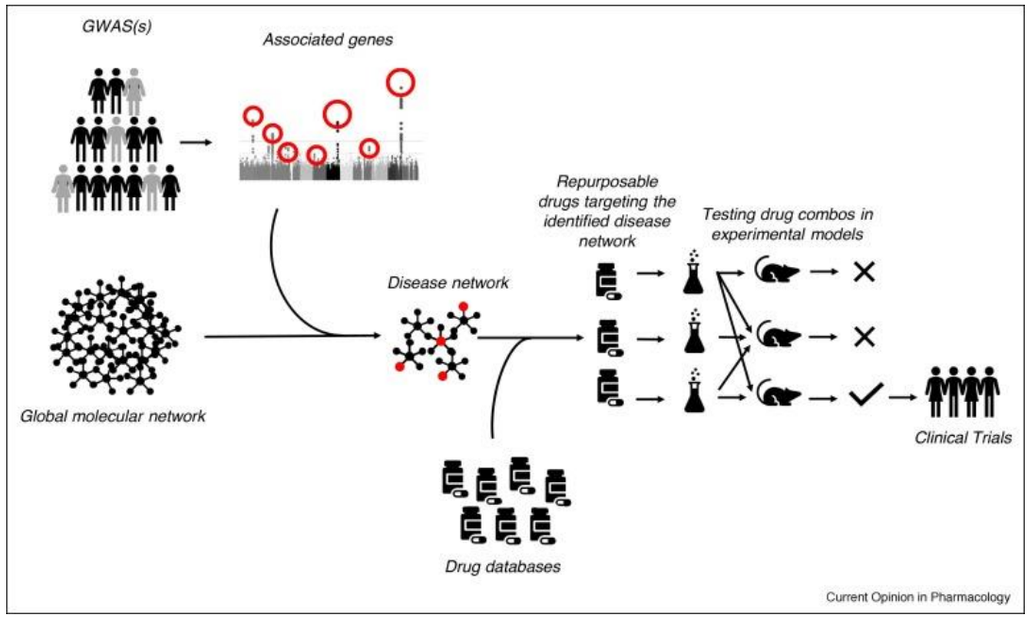


Table 1 | **Selected successful drug repurposing examples and the repurposing approach employed**

Drug name	Original indication	New indication	Date of approval	Repurposing approach used	Comments on outcome of repurposing
Zidovudine	Cancer	HIV/AIDS	1987	In vitro screening of compound libraries	Zidovudine was the first anti-HIV drug to be approved by the FDA
Minoxidil	Hypertension	Hair loss	1988	Retrospective clinical analysis (identification of hair growth as an adverse effect)	Global sales for minoxidil were US$860 million in 2016 (Questale minoxidil sales report 2017; see Related links)
Sildenafil	Angina	Erectile dysfunction	1998	Retrospective clinical analysis	Marketed as Viagra, sildenafil became the leading product in the erectile dysfunction drug market, with global sales in 2012 of $2.05 billion[8]
Aspirin	Analgesia	Colorectal cancer	2015	Retrospective clinical and pharmacological analysis	US Preventive Services Task Force released draft recommendations in September 2015 regarding the use of aspirin to help prevent cardiovascular disease and colorectal cancer[52]

ADHD, attention deficit hyperactivity disorder; EMA, European Medicines Agency; FDA, US Food and Drug Administration; MS, multiple sclerosis; SUI, stress urinary incontinence.

Source: Pushpakom et al. (2019)

Background: conventional approaches to drug repurposing

- Conventional computational and experimental repurposing methods focuses on pre-clinical outcomes and might not be consistent with therapeutic effects in humans
  - Structural features of compounds or proteins
  - GWAS
- Novel data sources such as electronic health record (HER)

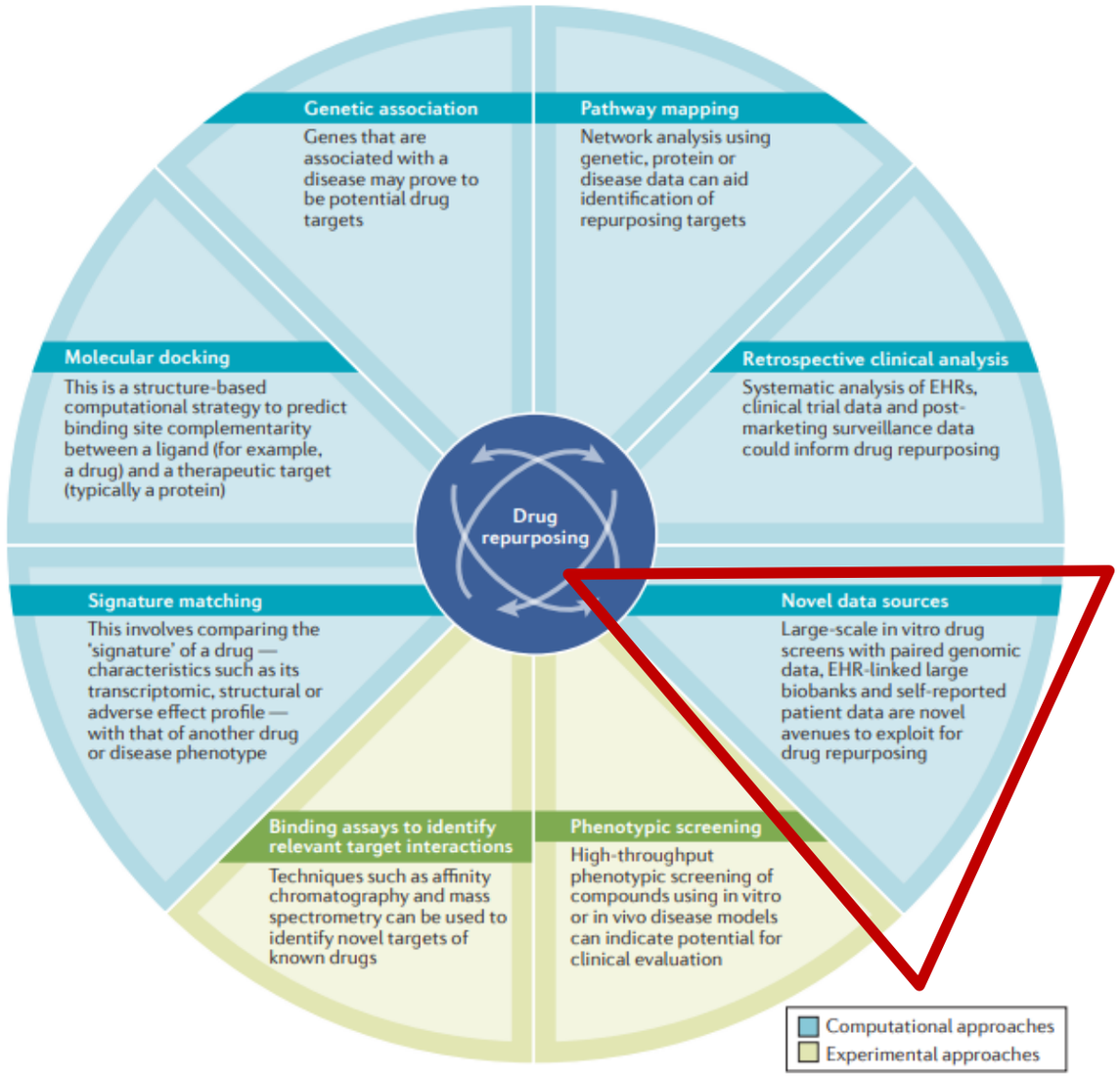
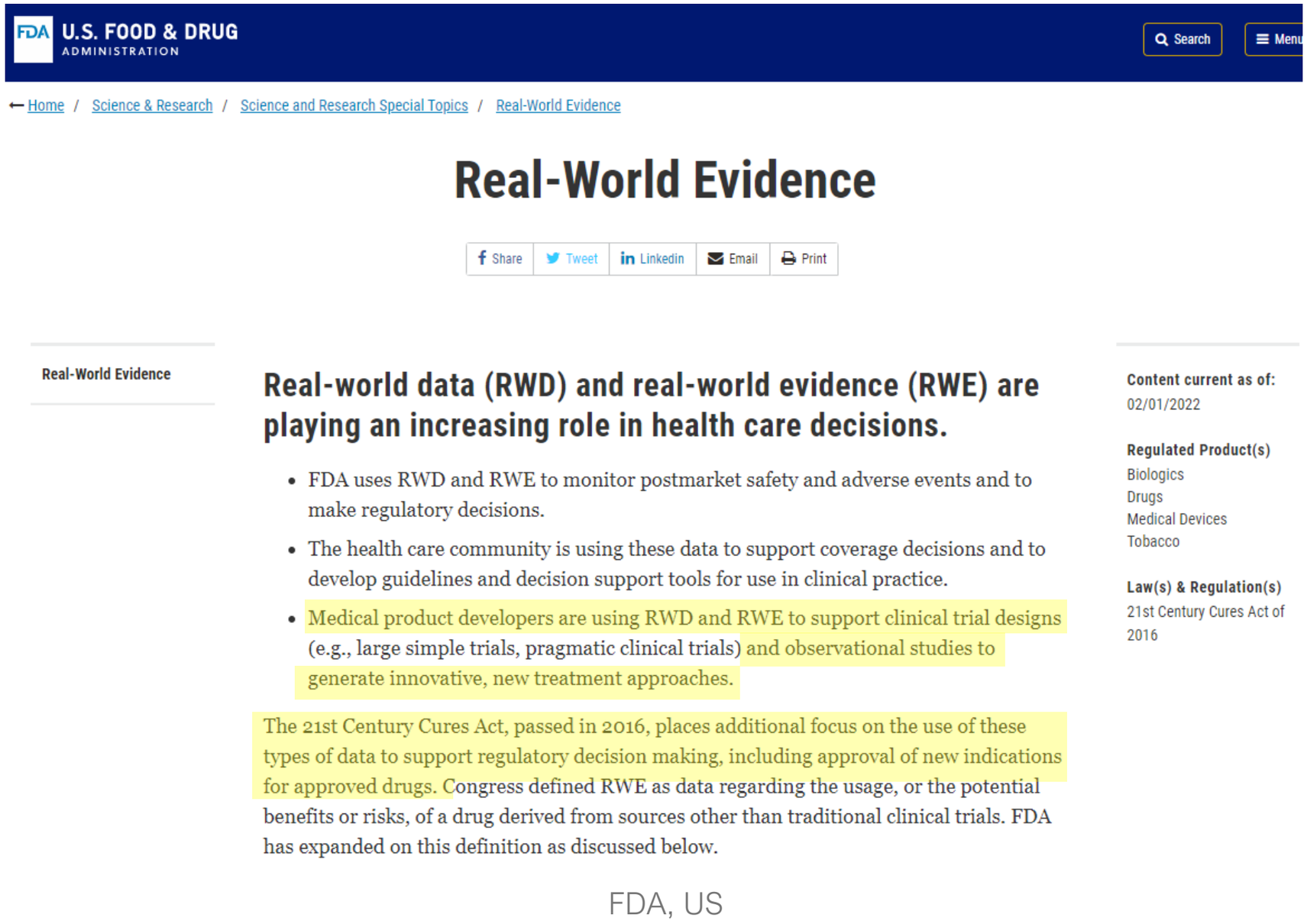


Figure 1 | **Approaches used in drug repurposing.** Various computational approaches can be used individually or in combination to systematically analyse different types of large-scale data to obtain meaningful interpretations for repurposing hypotheses. Challenges for such analyses are discussed in BOX 5. Experimental approaches can also be used to identify repurposing opportunities. EHR, electronic health record.

Source: Pushpakom et al. (2019)

Background: regulatory incentives



FDA U.S. FOOD & DRUG ADMINISTRATION

Search Menu

← Home / Science & Research / Science and Research Special Topics / Real-World Evidence

Real-World Evidence

Share Tweet Linkedin Email Print

Real-World Evidence

Real-world data (RWD) and real-world evidence (RWE) are playing an increasing role in health care decisions.

- FDA uses RWD and RWE to monitor postmarket safety and adverse events and to make regulatory decisions.
- The health care community is using these data to support coverage decisions and to develop guidelines and decision support tools for use in clinical practice.
- Medical product developers are using RWD and RWE to support clinical trial designs (e.g., large simple trials, pragmatic clinical trials) and observational studies to generate innovative, new treatment approaches.

The 21st Century Cures Act, passed in 2016, places additional focus on the use of these types of data to support regulatory decision making, including approval of new indications for approved drugs. Congress defined RWE as data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials. FDA has expanded on this definition as discussed below.

Content current as of:
02/01/2022

Regulated Product(s)
Biologics
Drugs
Medical Devices
Tobacco

Law(s) & Regulation(s)
21st Century Cures Act of 2016

FDA, US

Background: regulatory incentives

GUIDANCE DOCUMENT

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Draft Guidance for Industry

SEPTEMBER 2021

Download the Draft Guidance Document | Read the Federal Register Notice

Draft | Level 1 Guidance

Not for implementation. Contains non-binding recommendations.

This guidance complements the 2013 guidance by expanding on certain aspects of that guidance relating to the selection of data sources and also provides additional guidance for evaluating the relevance and reliability of both EHRs and medical claims data for use in a clinical study. This guidance also provides a broader overview of considerations relating to the use of EHR and medical claims data in clinical studies more generally, including studies intended to inform FDA's evaluation of product effectiveness.

FDA, US

Background: regulatory incentives

国家药品监督管理局
National Medical Products Administration

中 En

请输入关键字

国家药监局关于发布真实世界证据支持药物研发与审评的指导原则（试行）的通告(2020年第1号)

发布时间：2020-01-07

为进一步指导和规范真实世界证据用于支持药物研发和审评的有关工作，保障药物研发工作质量和效率，国家药品监督管理局组织制定了《真实世界证据支持药物研发与审评的指导原则（试行）》，现予发布。

特此通告。

附件：1.真实世界证据支持药物研发与审评的指导原则（试行）

2.《真实世界证据支持药物研发与审评的指导原则（试行）》的起草说明

Background: existing studies

- Hernán and Robins (2016) outlined a framework to **emulate randomised controlled trials (RCTs)** and assess treatment effects using observational data.

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

- Drug repurposing + observational data research by IBM Research: Ozery-Flato et al. (2020)

Research and Applications

Framework for identifying drug repurposing candidates from observational healthcare data

Michal Ozery-Flato [1], Yaara Goldschmidt[2,†], Oded Shaham[2,‡], Sivan Ravid[1], and Chen Yanover [2,§]

[1]Healthcare Informatics, IBM Research-Haifa, Mount Carmel Haifa, Israel and [2]Formerly Healthcare Informatics, IBM Research-Haifa, Mount Carmel Haifa, Israel

[†]Present address: K Health, Tel Aviv, Israel.
[‡]Present address: MeMed Dx, Haifa, Israel.
[§]Present address: KI Research Institute, Kfar Malal, Israel.
Corresponding Author: Michal Ozery-Flato, PhD, IBM Research-Haifa, Haifa University Campus, Mount Carmel Haifa, 3498825, Israel; ozery@il.ibm.com

Received 19 May 2020; Revised 11 September 2020; Editorial Decision 13 September 2020; Accepted 17 September 2020

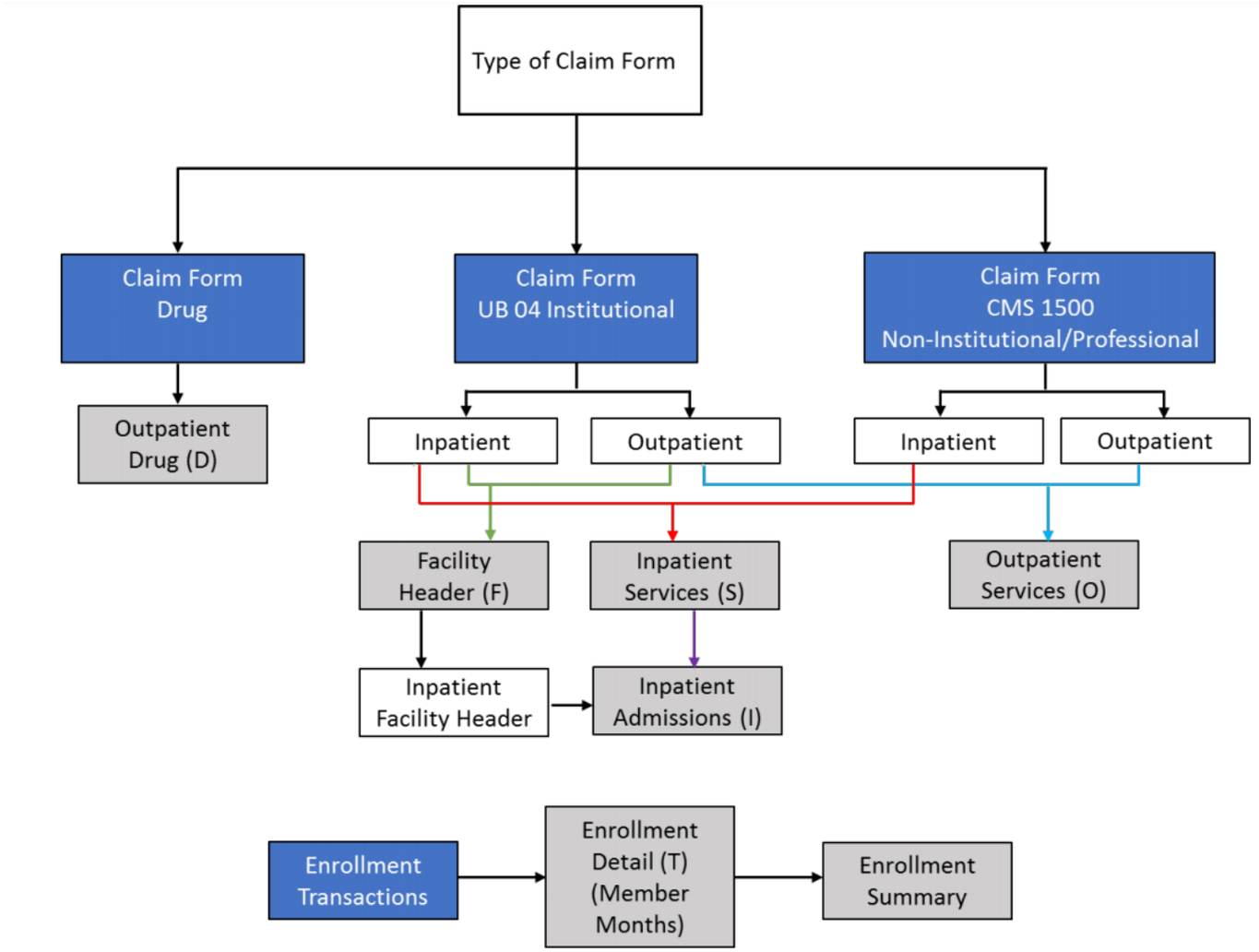
Study aim

- To develop a deep-learning based framework that emulates the RCTs for drugs present in a large-scale medical claims database to identify drug repurposing candidates with no previous indication on the disease of interest.

- The paper demonstrated a test case in a **coronary artery disease (CAD)** cohort.

Method: dataset

- *IBM® MarketScan®* (MarketScan) data contains individual-level, de-identified healthcare claims data in the US. The dataset recorded 1,178,997 CAD patients between 2012 and 2017.
- 3 source tables used are outpatient drug, inpatient admission, outpatient services.
- 3 types of study variables:
  - Demographic characteristics: age and gender
  - Disease outcomes ascertained by ICD9/ICD10 codes.
  - Drugs prescribed: 1,353 unique drug ingredients.



Method: data pre-processing

MarketScan database

Disease outcomes
Ascertained by ICD9/ICD10 codes.

Demographics
Age and gender

Drugs prescribed*
1,353 unique drug ingredients.

CAD patients
1,178,997 patients

Hypothesis
Effective drugs would lower the risk of CAD patients developing those diseases / events.

Prescribed drug candiates
55 drug ingredient / 38 drug-class candidates / 7 drug combinations **
*** remained as **RCT emulator**.

Study outcome
Heart failure or stroke after diagnosis of CAD

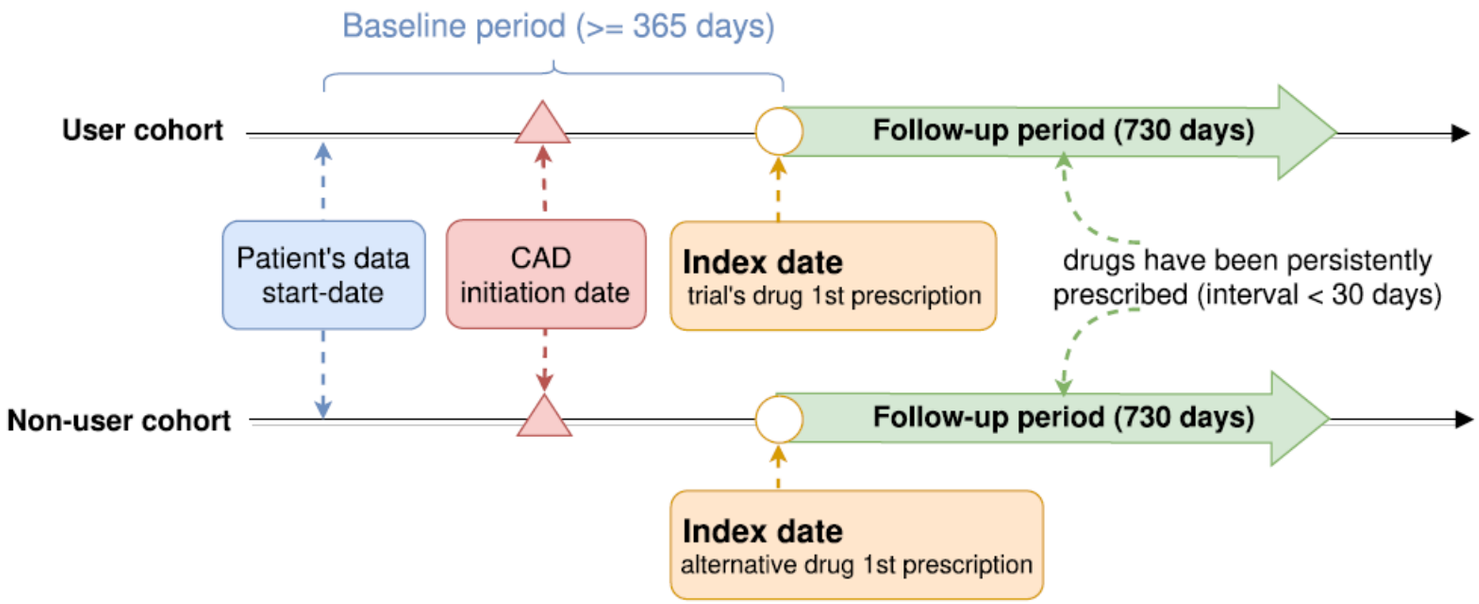
User and non-user cohorts

Confounders
Age (0-1), gender (binary), co-morbidities and co-prescribed drugs

Domain expert

* For drugs with multiple ingredients, the individual active ingredients were considered separately.
** This study considered both drug ingredients, drug-class and drug combinations of 2 as repurposing emulators, however results in the main report were reported by drug ingredient.
*** A set of conditions needs to be met for a drug / drug class to be considered repurposing candidates, see pg.73 for details.

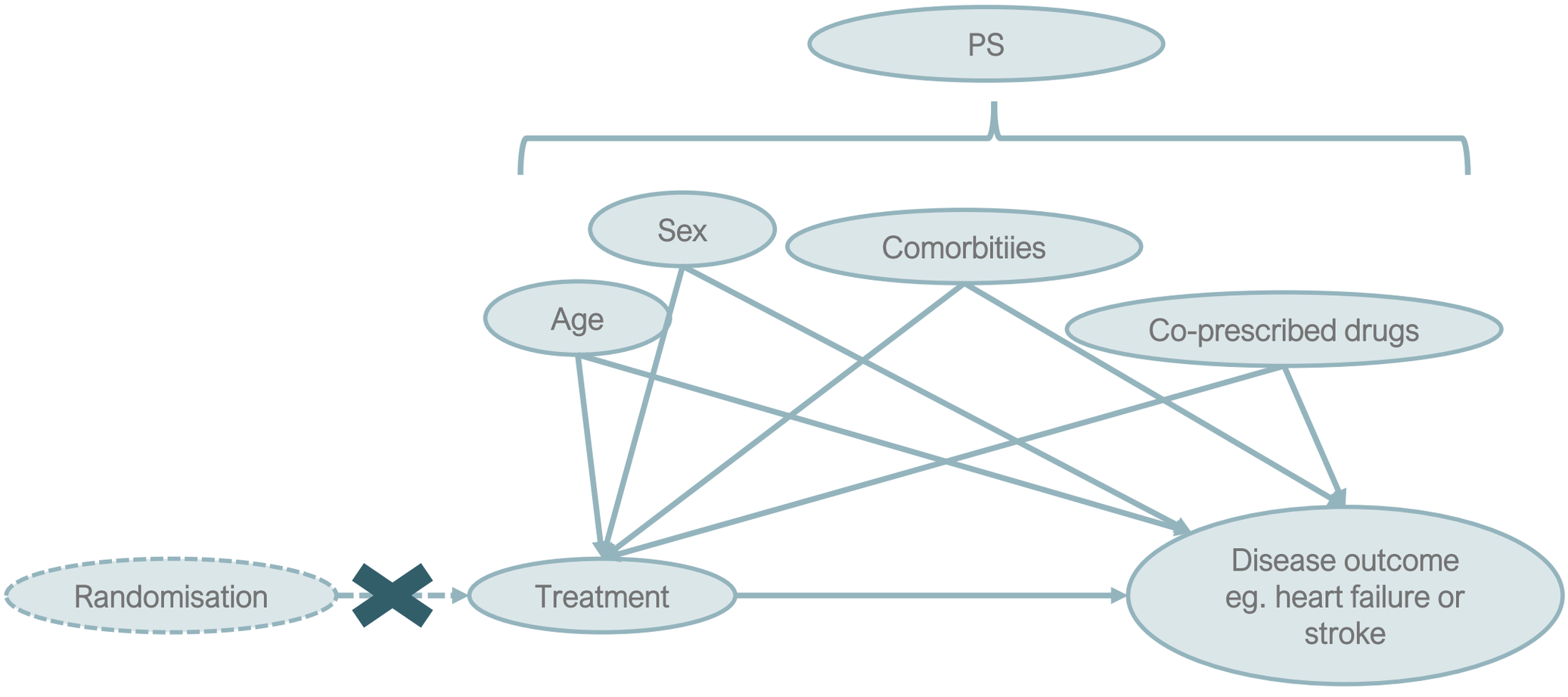
Method: definition of user and non-user cohorts



Extended Data Fig. 8 | The definition of user and non-user cohorts. Index date refers to the first prescription of the trial's drug (user cohort) or the alternative drug (non-user cohort). The time period before the index date is the baseline period, and the time after the index date is the follow-up period. The patient covariates are collected during the baseline period and the treatment effects areevaluated at the follow-up period.

Method: model for propensity score (PS) weighting

- Logistic regression-based (LR-based) PS weighting as **baseline**. The aim of the model is to predict treatment assignment.



Method: model for propensity score (PS) weighting

- What is an LSTM?
  - A type of **recurrent neural network** that allows information to persist (over time).
  - Can be used to make predictions based on **time series data**.

- LSTM-based PS estimation for treatment assignment:
  - Embedding module
  - LSTM module with attention mechanism
  - Prediction module

- Loss function: **binary cross entropy** where the target values are treatment assignment {0,1}.

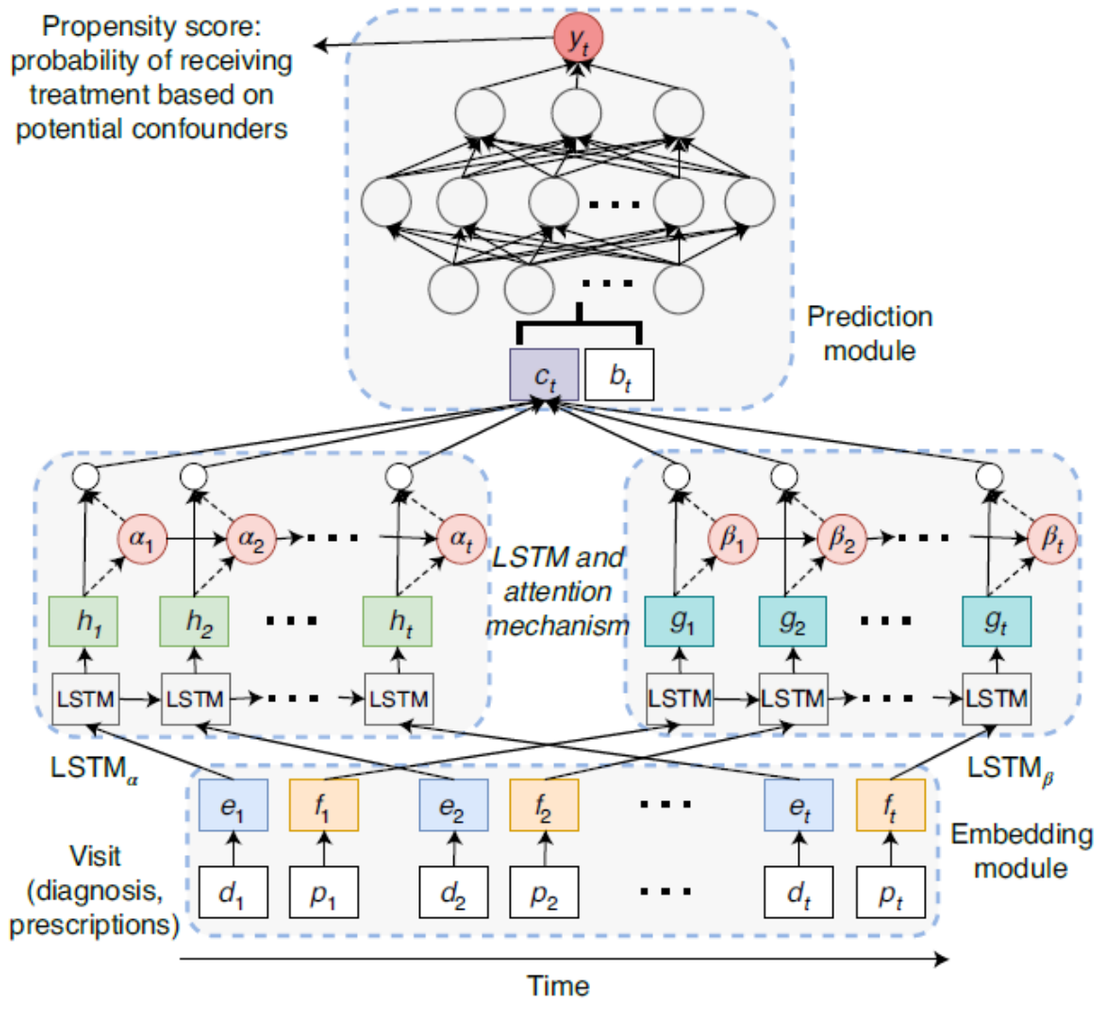


Fig. 2 | Illustration of the deep learning model for predicting treatment probability (or propensity score) that we used to correct confounding from time sequence data (including diagnoses d_t, prescriptions p_t and demographics b_t). It consists of three main components: an embedding module, a recurrent neural network (LSTM) and a prediction module.

Method: model for propensity score (PS) weighting

- The embedding module includes 2 linear embedding layers for diagnosis and prescription respectively.

d_1
p_1

$\{d_1, d_2, \dots, d_t\}$ = diagnosis code for each patient at each timestamp, $\{p_1, p_2, \dots, p_t\}$ = prescription. We have $dim(\boldsymbol{d_t}) = 1 \times r$ and $dim(\boldsymbol{p_t}) = 1 \times s$, where r = size of the diagnosis dictionary and s= size of the prescription code dictionary.

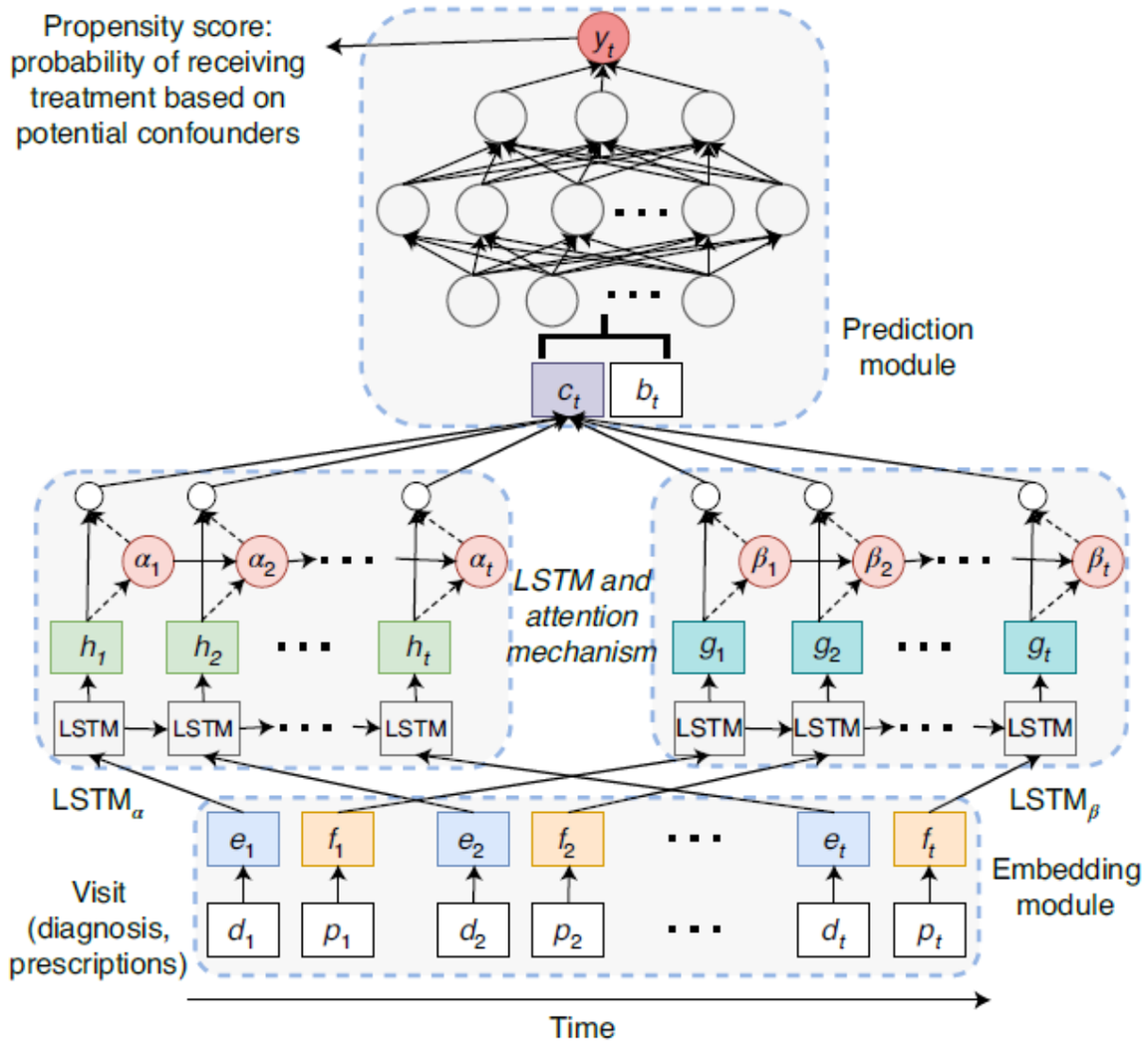
e_1
d_1

$\boldsymbol{e}_t = \boldsymbol{W}^d_{emb}\boldsymbol{d}_t \in R^m$ denotes the embedding of the input vector $\boldsymbol{d}_t \in R^r$, where m is the size of the **diagnosis** embedding dimension.

f_1
p_1

$\boldsymbol{f}_t = \boldsymbol{W}^p_{emb}\boldsymbol{P}_t \in R^n$ denotes the embedding of the input vector vector $\boldsymbol{p}_t \in R^s$, where n is the size of the **prescription** embedding dimension.

$\boldsymbol{W}^d_{emb} \in R^{m \times r}$ and $\mathbf{W}^p_{emb} \in R^{n \times s}$ denote the embedding matrices of **diagnosis** and **prescription** respectively.



Method: model for propensity score (PS) weighting

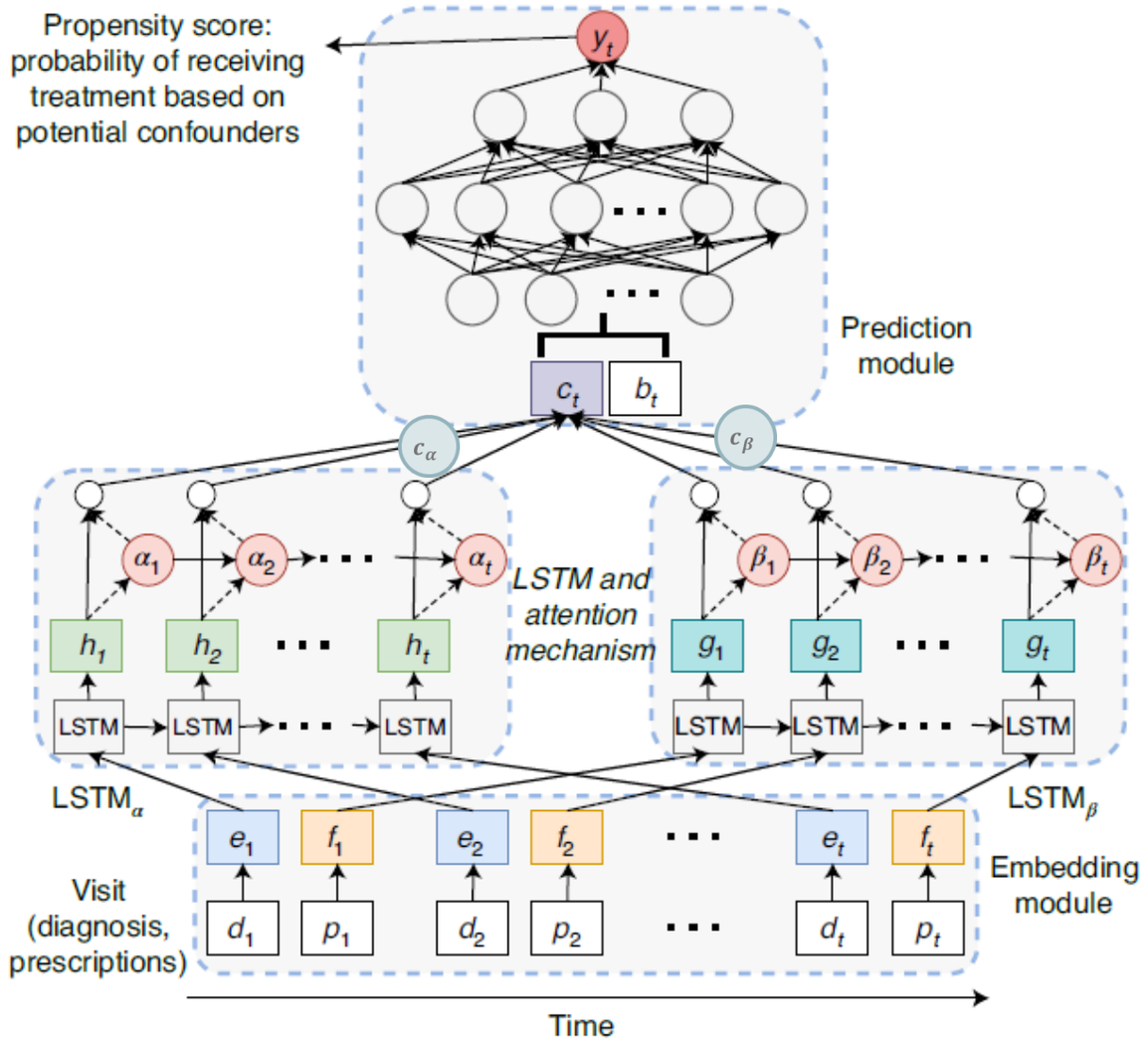
- The LSTM module with attention mechanism can model the temporality of observational data. Two bi-directional LSTM layers LSTM_α and LSTM_β to model diagnosis and prescription of participants.

$$\begin{aligned} \mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_t &= \mathrm{LSTM}_\alpha(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_t) \\ \mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_t &= \mathrm{LSTM}_\beta(\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_t) \end{aligned} \tag{3}$$

where $\mathbf{h}_t \in \mathbb{R}^u$, $\mathbf{g}_t \in \mathbb{R}^v$ are hidden state vectors at t-th visit, and u and v denote the size of hidden layer of LSTM_α and LSTM_β. Then those patient hidden states are aggregated through two separate attention layers for automatically focusing on important visits.

$$\begin{aligned} \alpha_i &= \mathrm{Softmax}\left(\mathbf{W}_\alpha^\top \mathbf{h}_i + \mathbf{b}_\alpha\right), \quad \text{for } i = 1, 2..., t \\ \mathbf{c}_\alpha &= \sum_{i=1}^{t} \alpha_i \mathbf{h}_i \\ \beta_i &= \mathrm{Softmax}\left(\mathbf{W}_\beta^\top \mathbf{g}_i + \mathbf{b}_\beta\right), \quad \text{for } i = 1, 2..., t \\ \mathbf{c}_\beta &= \sum_{i=1}^{t} \beta i \mathbf{g}_i \end{aligned} \tag{4}$$

where $\mathbf{W}_\alpha \in \mathbb{R}^u$, $\mathbf{b}_\alpha \in \mathbb{R}^u$, $\mathbf{W}_\beta \in \mathbb{R}^v$ and $\mathbf{b}_\beta \in \mathbb{R}^v$ are the parameters to learn. Using the generated attention weights for diagnosis and prescription, we obtain the aggregated vectors $\mathbf{c}_\alpha \in \mathbb{R}^u$ and $\mathbf{c}_\beta \in \mathbb{R}^v$ as defined in equation (4). Then we combine $\mathbf{c}_\alpha$, $\mathbf{c}_\beta$ with vectorized age and gender to predict the probability of receiving a treatment (propensity score).



Method: model for propensity score (PS) weighting

- Prediction module: predicts the propensity score based on patient states from the attention layer in the LSTM module (diagnosis, prescription), as well as patient age and sex.

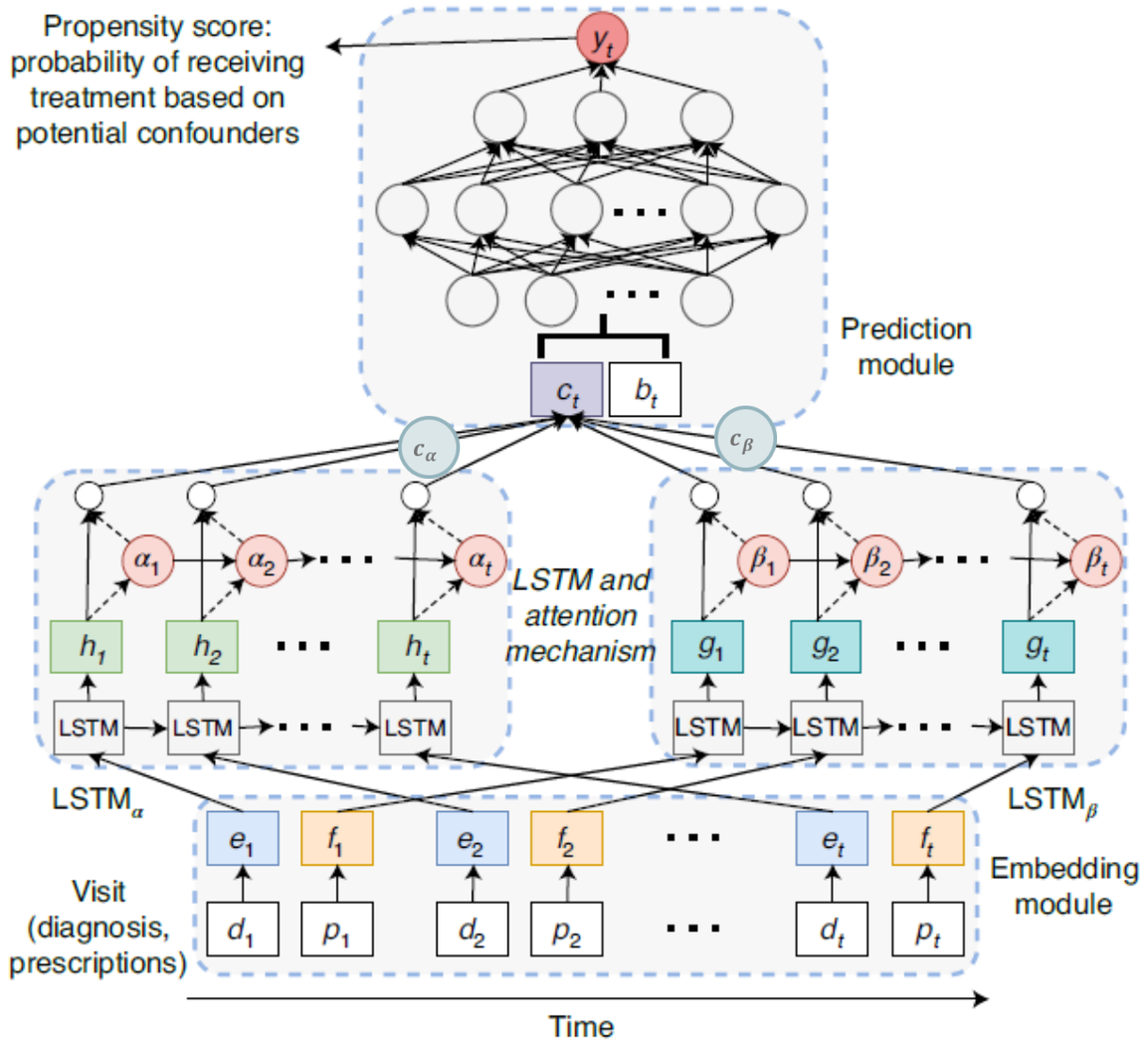
Prediction module. The aggregated patient states from attention layer $\mathbf{c}_\alpha$, $\mathbf{c}_\beta$, combined with the demographic features c_{demo}, are passed through a fully connected neural network to predict the probability of receiving a treatment as follows,

$$\hat{y} = \text{Sigmoid}\left(\mathbf{W}^\top \mathbf{c}_t + b\right) \tag{5}$$

where $\mathbf{c}_t = \text{ReLu}(\mathbf{W}_c[\mathbf{c}_\alpha, \mathbf{c}_\beta, c_{\text{demo}}] + \mathbf{b}_c)$, $\mathbf{W}_c \in \mathbb{R}^{k \times (u+v+2)}$, $\mathbf{b}_c \in \mathbb{R}^k$, $\mathbf{W} \in \mathbb{R}^k$, $b \in \mathbb{R}$ are the model parameters. We use cross-entropy to calculate the prediction loss as follows,

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log \hat{y}_i + (1-y_i)\log\left(1-\hat{y}_i\right)\right) \tag{6}$$

where y_i is the ground truth of observed treatment for patient i.



Methods: standardised inverse probability weighting (IPTW)

- Conventional IPTW approach:
  - The weight of a user is calculated as $w(x) = \frac{1}{p(x)}$, the weight of a non-user is $w(x) = \frac{1}{1-p(x)}$.
  - Here $p(x)$ denotes the propensity score.
  - However, one common issue in IPTW is that some participants may receive:
    - Very low PS → very large weight
    - Very high PS → very small weight

- To address the issue, an alternative weighting approach called **standardised IPTW** (Austin and Stuart, 2015) were adopted:
  - That is, the weight of a user is calculated as $w(x) = \frac{P(T=1)}{P(T=1 \mid X=x)}$, while that for a non-user is $w(x) = \frac{1-P(T=1)}{1-P(T=1 \mid X=x)}$.
  - Here T denotes the treatment assignment.

- Extreme weights were subsequently clipped at 1e-06 and 100.

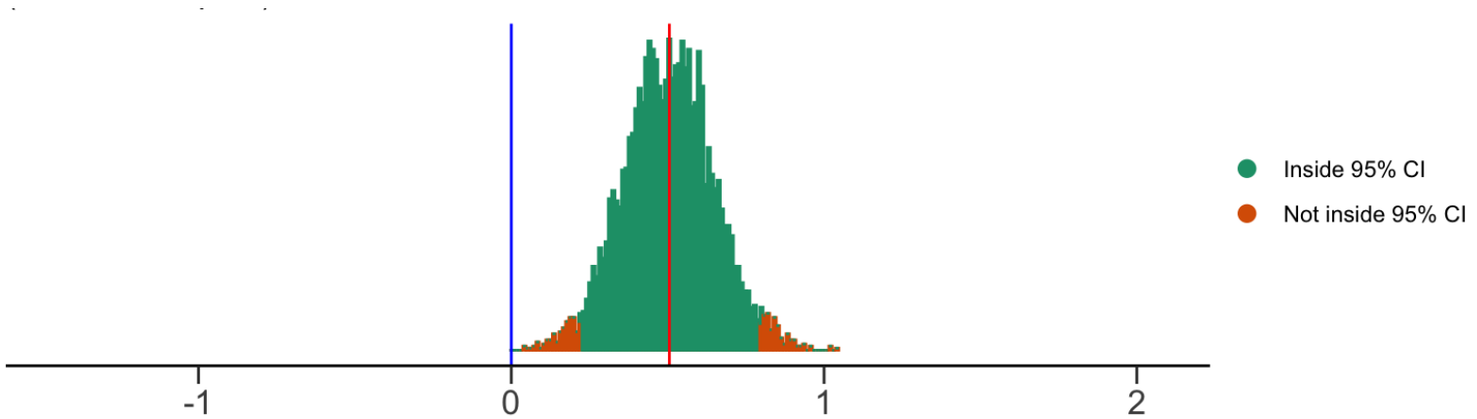
Method: average treatment effect (ATE)

- For each candidate drug, its **ATE** is calculated at the population-level to see it can improve the disease outcomes. ATE is defined as:

$$\mathrm{ATE} = \mathbb{E}(Y_1) - \mathbb{E}(Y_0) \qquad (1)$$

where $\mathbb{E}(Y_1)$ and $\mathbb{E}(Y_0)$ are the expected potential treated and control outcomes of the whole population, respectively.

- **95% confidence interval (CI) and p-values** are calculated using **bootstrapping** (e.g. 50 times with replacement).
  - 95% CI: mean ± 1.96 × standard error
  - P-value: normal cumulative distribution function of estimators.
  - Adjusted p-value for multiple testing (<0.05) (Benjamini and Hochberg, 1995)



Method: model performance / feature balancing

- Once the weighted user and non-user sub-cohorts are generated by IPTW, the balance of confounders are quantified using the **standardised mean difference (SMD)** between the two treatment groups. SMD is defined as follows:

$$\mathrm{SMD} = \frac{|\mu_{\mathrm{user}} - \mu_{\mathrm{non-user}}|}{\sqrt{({s_{\mathrm{user}}}^2 + {s_{\mathrm{non-user}}}^2)/2}} \qquad (2)$$

where μ_{user} and $\mu_{\mathrm{non\text{-}user}}$ are the mean in user cohort and non-user cohort; ${s_{\mathrm{user}}}^2$ and ${s_{\mathrm{non-user}}}^2$ are sample variance of variables in two sub-cohorts.

- For each confounder, when its SMD > 0.1 it may be considered **unbalanced** between the treatment groups.
- Across all confounders, the **unbalanced feature ratio** (unbalanced / all features) before and after weighting should be calculated to evaluate the performance of balancing**.** If the unbalanced feature ratio is **below 2%** after weighting, the user and non-user sub-cohorts could still be considered as **balanced**.
- The learned **attention weights** were used to visualise each confounder and its SMD values before/after balancing, between the user and non-user cohorts.

Method: model training

Training / test / validation split	70:10:20

Hyperparameter	Details
Batch size	50
Random seed	128
Learning rate	1e-03
Weight decay	1e-06
Number of epochs	50 epochs

Method: pipeline

Input: patient data: assigned treatment, outcomes, values for potential confounders

Output: repurposed drug candidates, and their estimated effect, unbalanced feature ratio and significance

1: Generate user and non-user sub-cohorts for the treatment
2: Compute balancing weights for all patients in both sub-cohorts via LSTM-based IPTW
3: Estimate the effect over multiple outcomes after correcting for the biases in the confounders (equation (1))
4: Compute the unbalanced feature ratio for the treatment after re-weighting using standardized difference (equation (2))
5: Estimate the significance of effect and compute adjusted p-values using bootstrapping
6: **if** estimated effect < 0 and adjusted p-value < 0.05 and unbalanced feature ratio $< 2\%$ **then**
7: **return** the estimated effect, unbalanced feature ratio and computed p-value
8: **end if**

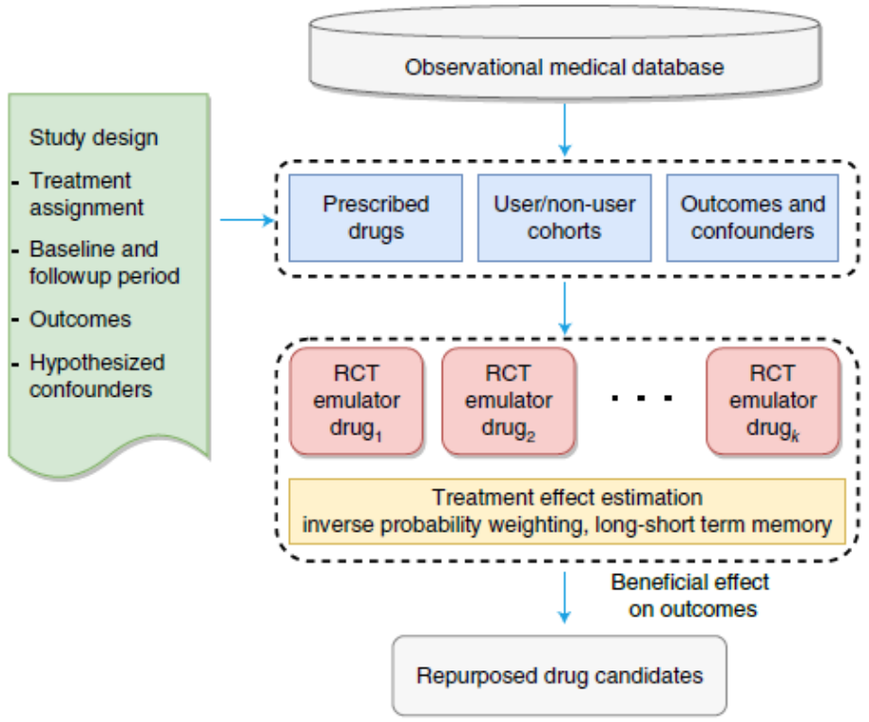
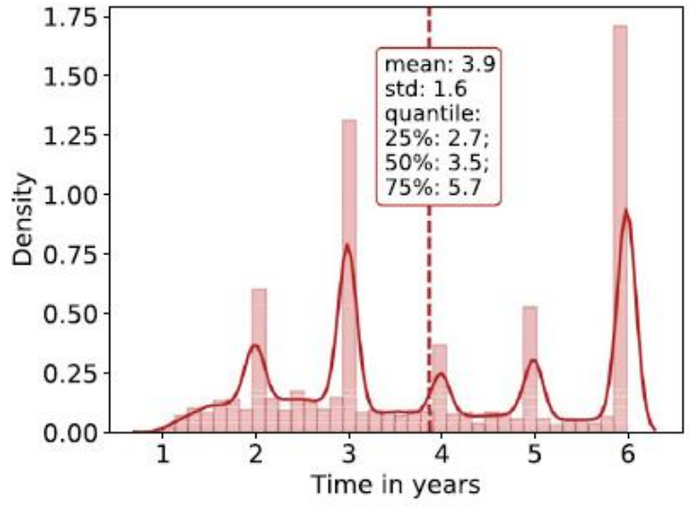


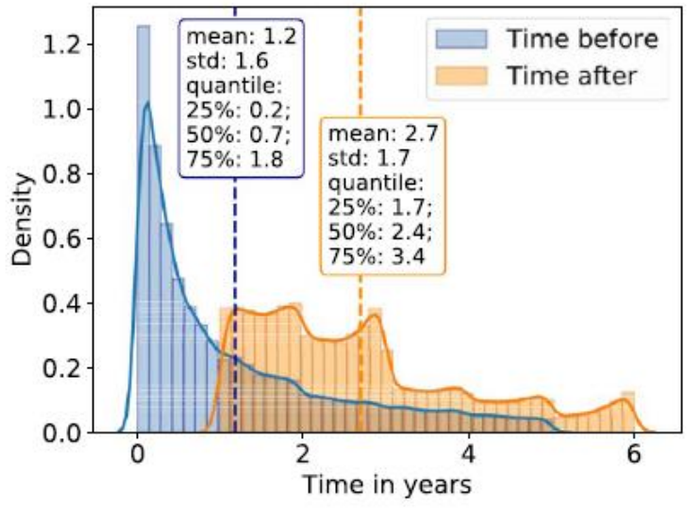
Fig. 1 | Flowchart of overall drug repurposing framework. First, a list of potential repurposing drug ingredients are extracted from the observational medical database given a disease cohort. Second, for each ingredient, the framework identifies the corresponding user and non-user sub-cohorts, and computes a large number of features for patients in both sub-cohorts. Third, the treatment effects are estimated via emulating an RCT for each ingredient to adjust confounding and biases.

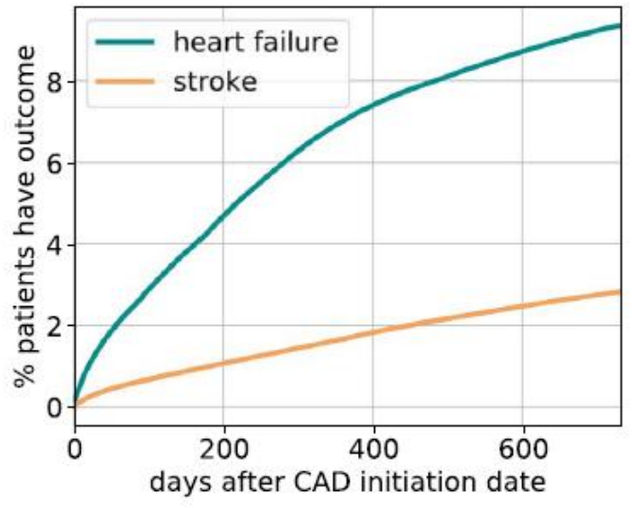
Method: experiments

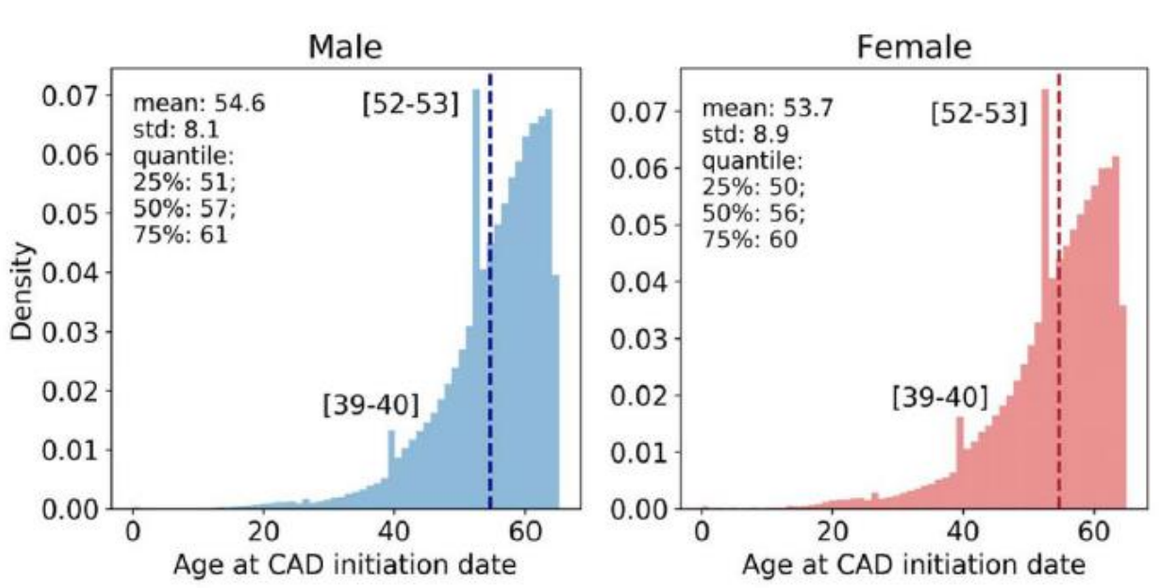
Experiments	Details	Evaluation metric
Comparison between LSTM-based and LR-based methods		AUC and feature balancing.
Comparison between drug ingredients and drug classes	55 drug ingredients and 38 drug classes.	P-values and post-balancing ATE.
Experiments on drug combinations	7 drug combination as repurposing candidates.	P-values and post-balancing ATE.
Comparison with pre-clinical-based methods	Experiments were conducted using three pre-clinical drug repurposing methods for CAD: • Chemical-protein interactome docking (CPI) • Chemical structures • Drug targets	Precision@*k* were used as evaluation metric, where *k* represents the k top-ranked drug candidates. This is due to limited budget at pharmaceutical companies.
Influence of adjusted p-value	Relaxing the threshold for adjusted p-value from 0.05 to 0.15.	

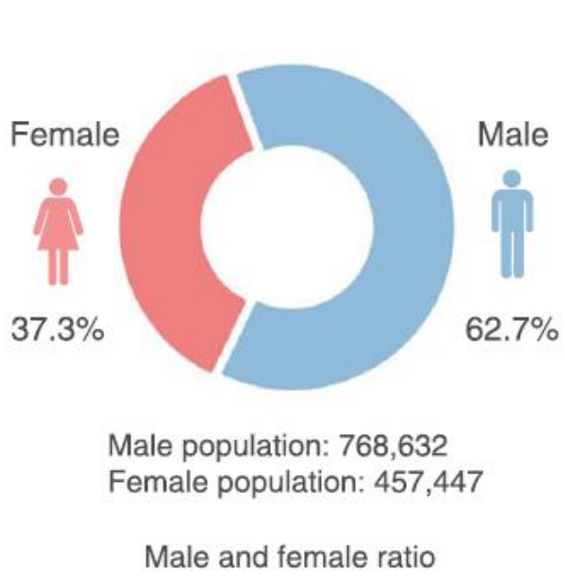
Results: CAD cohort characteristics











Extended Data Fig. 1 | CAD cohorts characteristics. a, The patients' distribution of total time in the database. **b**, The patient's distribution of time before/after CAD *initiation date*. **c**, The growth of the number of patients developing outcomes after CAD *initiation date*. **d**, The gender distribution with age at CAD *initiation date*.

Results: feature balancing and ATE estimates

Table 1 | The estimated treatment effects for CAD over balanced and statistically significant drug ingredients

Drug name	Users	Non-users	Unbalanced covariates (pre)	Unbalanced covariates (post)	Covariates	Unbalanced ratio (post)	ATE (pre)	ATE (post)
Metoprolol	**9,730**	**29,190**	**38.308**	**23.231**	**1,270**	**1.8**	**-0.023**	**-0.043**
Fenofibrate	**1,352**	**4,056**	**39.340**	**13.200**	**1,038**	**1.3**	**-0.051**	**-0.038**
Rosuvastatin	2,420	7,260	24.020	9.620	1,097	0.9	-0.063	-0.030
Hydrochlorothiazide	**2,001**	**6,003**	**32.500**	**15.320**	**1,076**	**1.4**	**-0.055**	**-0.029**
Amlodipine	4,613	13,839	21.340	8.300	1,180	0.7	-0.050	-0.026
Pravastatin	**2,007**	**6,021**	**11.260**	**9.640**	**1,085**	**0.9**	**-0.016**	**-0.022**
Simvastatin	**1,605**	**4,815**	**10.060**	**13.240**	**1,044**	**1.3**	**-0.032**	**-0.020**
Valsartan	**1,316**	**3,948**	**24.940**	**13.740**	**1,026**	**1.3**	**0.010**	**-0.015**
Diltiazem	1,044	3,132	28.360	13.080	1,007	1.3	-0.010	-0.013
Isosorbide	1,482	4,446	33.320	9.560	1,039	0.9	0.045	0.034
Prasugrel	1,316	3,948	41.500	18.340	1,019	1.8	-0.043	0.036
Ramipril	887	2,661	25.340	14.840	973	1.5	0.020	0.043
Potassium chloride	1,110	3,330	43.460	20.240	1,016	2.0	0.169	0.090
Carvedilol	3,959	11,877	38.280	8.140	1,154	0.7	0.198	0.124
Furosemide	1,545	4,635	50.880	17.080	1,064	1.6	0.301	0.179
Spironolactone	1,292	3,876	70.620	12.920	1,034	1.3	0.393	0.190

Bold denotes ingredients without a known CAD indication (repurposed drug candidates). The drugs are ranked by the estimated ATE values. Pre and post refer to re-weighting.

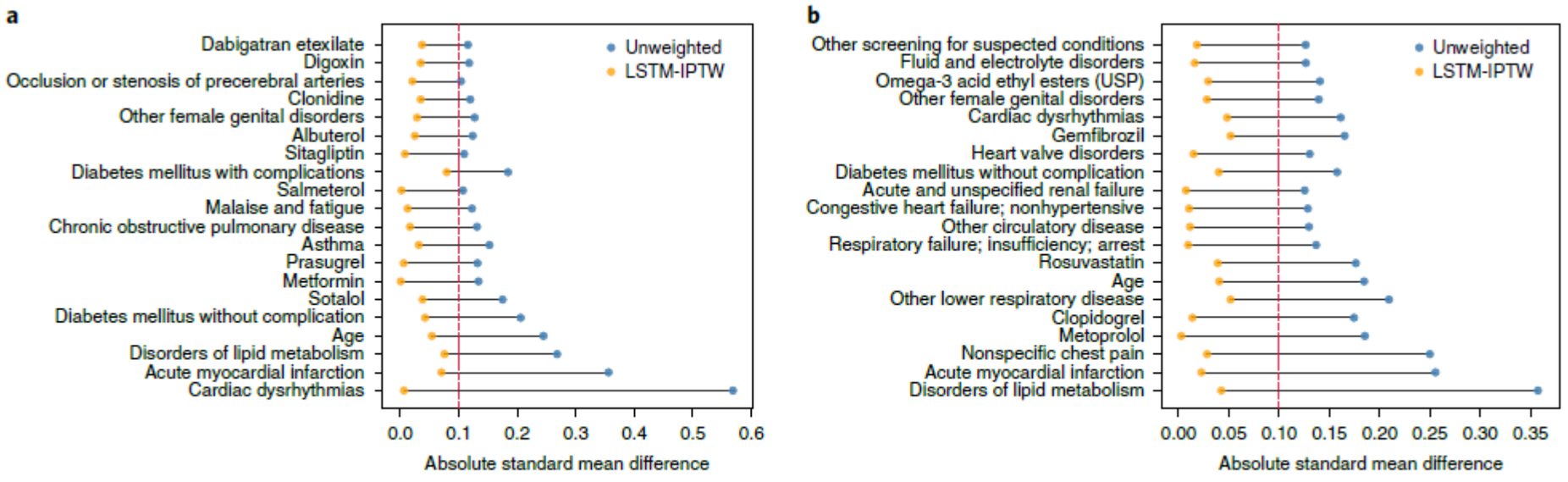


Fig. 4 | The SMD values of the top 20 well-balanced covariates. a, Diltiazem results. **b**, Fenofibrate results. The dashed red lines indicate the threshold of balancing.

Results: ATE estimates

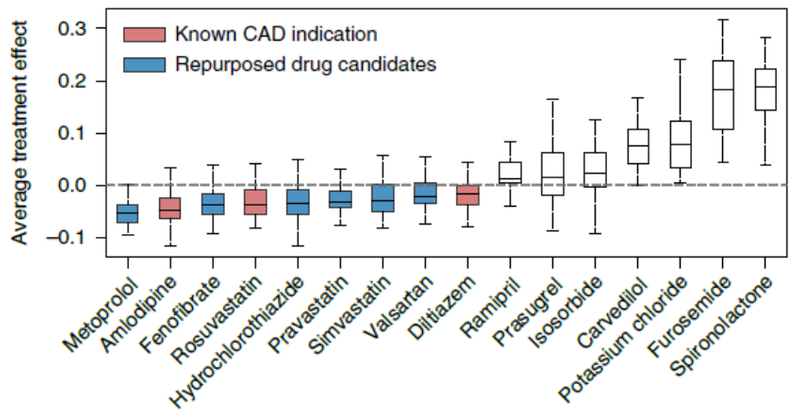
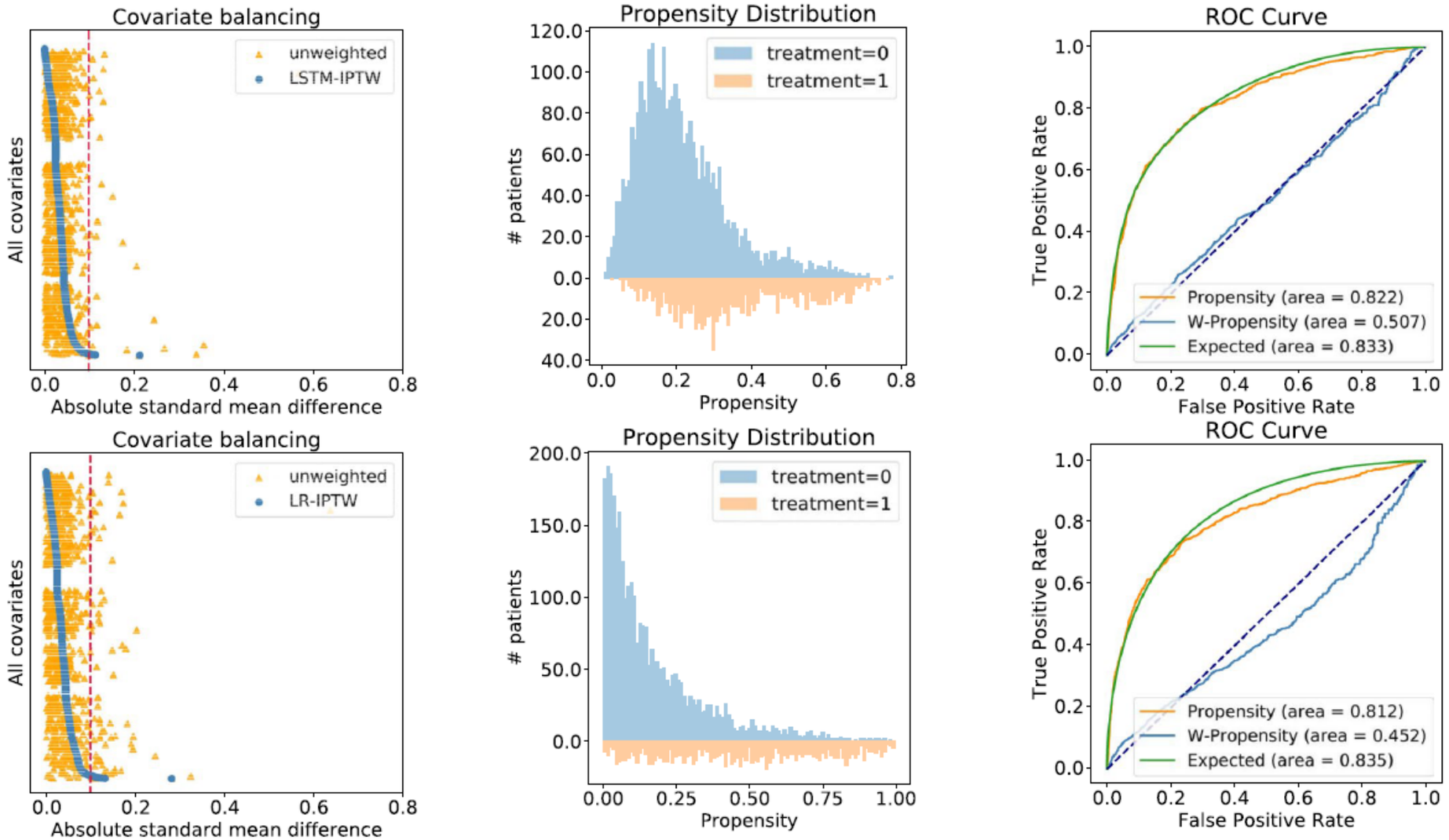


Fig. 3 | Distribution of estimated ATE of drugs on defined outcomes across the 50 bootstrap samples. All shown drugs satisfy two conditions: adjusted *p*-value ≤ 0.05 and post-weighting unbalanced ratio $\leq 2\%$. Within the boxplot, the central line denotes the median, and the bottom and the top edges denote the 25th (Q1) and 75th (Q3) and percentiles respectively. The whiskers extend to 1.5 times the interquartile range.

Results: LSTM-IPTW vs LR-IPTW



Extended Data Fig. 2 | Performance comparison of LSTM-IPTW and LR-IPTW using drug candidate: diltiazem (*with* known CAD indication). The three figures on the top are results obtained from LSTM-IPTW, while the figures on the bottom are from LR-IPTW. **a**, and (**d**) The absolute SMD of each covariate in the original data (orange triangles) and in the weighted data (blue circles). **b**, and (**e**) The distribution of estimated propensity scores over user (orange area) and non-user (blue area) cohorts. **c**, and (**f**) The ROC curves for the propensity model (orange), expected value (green) and weighted propensity (blue).

Results: drugs vs drug classes

- Comparison of results between drug ingredients and drug classes were **mostly consistent**, however there were **discrepancies**.
- Some examples are:
  - **Rosuvastatin**, **pravastatin** and **simvastatin** under drug class “HMG CoA reductase inhibitors” (ATC code: **C10AA**) showed significant result individually, but their drug class was not significant.
  - Some drug classes including “other antidepressants” and “selective serotonin reuptake inhibitors” showed a beneficial effect with statistical significance, but their drug ingredients were neither significance nor beneficial to the disease.

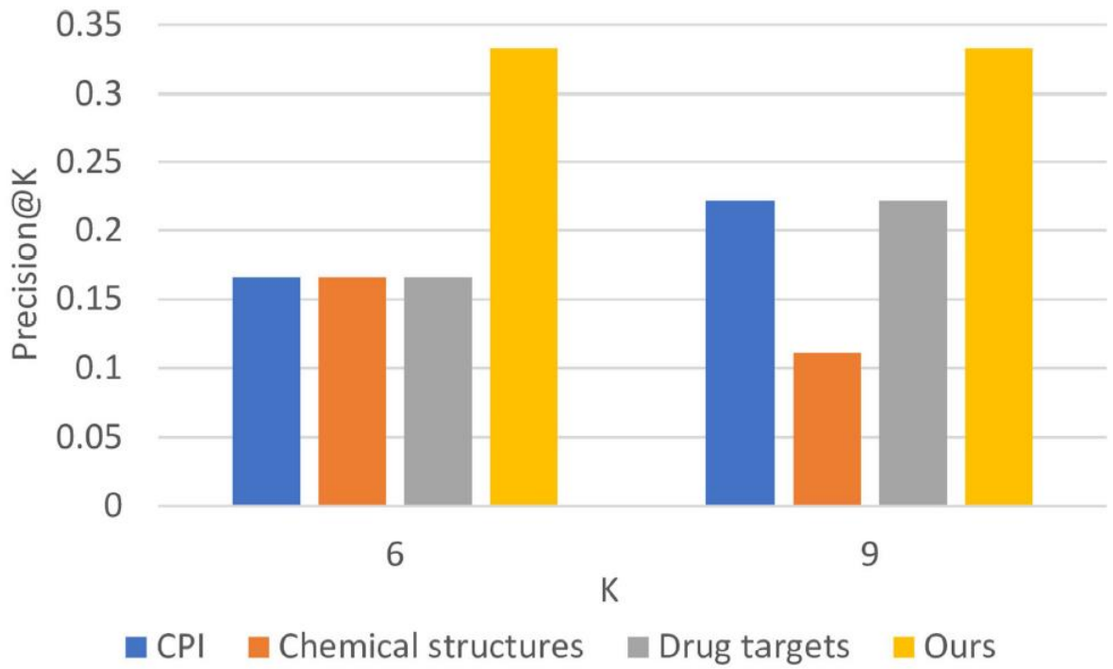
Results: drug combinations

- **Lisinopril** and **atorvastatin** were not statistically significant as individual drug ingredients, however their drug combinations were found to be both significant and beneficial to the disease outcomes.

Drug name	# User	# Non-user	Post.unbalanced.ratio%	Pre.ATE	Post.ATE	Adjusted p-value
Metoprolol + Clopidogrel	1237	3711	0.010	-0.034	-0.028	<0.05
Metoprolol + Atorvastatin	2158	6474	0.014	-0.045	-0.024	<0.05
Lisinopril + Atorvastatin	1145	3435	0.015	-0.002	-0.018	<0.05
Carvedilol + Atorvastatin	860	2580	0.011	0.124	0.112	<0.05

Extended Data Fig. 5 | The estimated treatment effects for CAD over balanced and statistically significant drug combinations. The drug combinations are ranked by the estimated ATE values.

Results: comparison with pre-clinical methods



Extended Data Fig. 6 | Performance comparison of proposed method and three pre-clinical methods evaluated by Precision@K. The values of K are selected from {6, 9}.

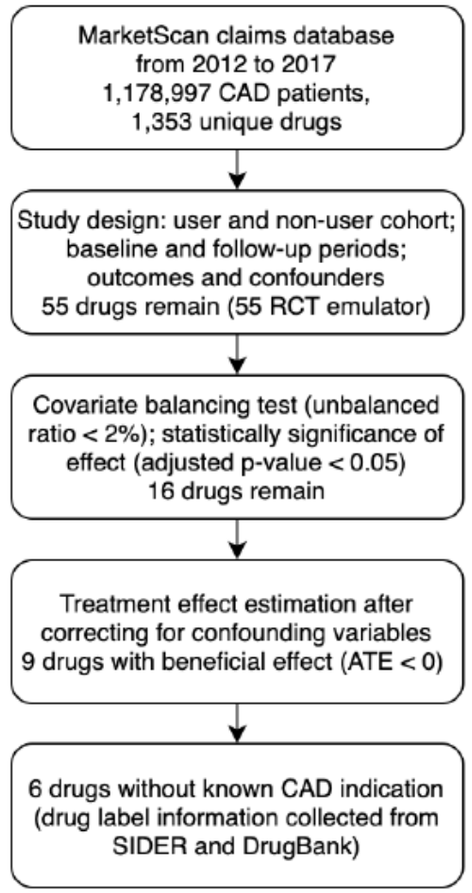
CPI = Chemical-protein interactome docking

Results: influence of adjusted p-value threshold

- 4 additional drug are found to be repurposing candidates when relaxing the adjusted p-value threshold from 0.05 to 0.15:
  - Metformin
  - Escitalopram*
  - Atorvastatin*
  - Losartan*
- The * drugs are already being assessed in studies or clinical trial(s) for CAD repurposing.

Discussions: main contributions and findings

- Demonstrated a deep-learning-based framework for high-throughput screening of on-market drugs by emulating, for each drug, an RCT that evaluates its treatment effect.
  - This allows **repurposed drug candidates** to be proactively generated from **existing large-scale RWD**.
  - Treatment effects were estimated using longitudinal observational data.
  - The framework was implemented in a CAD cohort that was automatically derived.
  - The proposed LSTM-PS estimation model could correct for confounding and selection bias, and more superior than the LR-regression-based PS estimation model.
- **Drug-class and drug combinations** could provide interesting findings of potential purposing.
- The proposed framework **more superior than pre-clinical methods**.



Supplemental Figure 2. Flowchart of data collection and study process of identifying repurposed drug candidates

Discussions: strengths and limitations

- The attention mechanism aided the interpretability and comparison of feature balancing.
- Taken into account multiple testing.
- Race and ethnicity not included as covariates.
- Only baseline confounders (age, sex, co-morbidities and co-prescribed drugs) are considered to inform the calculation of propensity scores. However, given that the study dataset is longitudinal, it bears the possibility that the confounders might have time-varying effects, eg. new co-morbidities, new co-prescribed drugs (within the scope by the input data).
- We know the baseline model choice of logistic regression is inherently unable to model longitudinal data effectively, whereas LSTM has the capability to capture time-varying confounders as input.

Future work

- Other potential confounders may be included in the model going forward:
  - Time elapsed from the first disease diagnosis to index date
  - Outcome value calculated over the baseline period

GitHub repo

https://github.com/ruoqi-liu/DeepIPW

References

- Hernán, Miguel A., and James M. Robins. 'Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available: Table 1.' American Journal of Epidemiology 183, no. 8 (15 April 2016): 758–64. https://doi.org/10.1093/aje/kwv254.
- Ozery-Flato, Michal, Yaara Goldschmidt, Oded Shaham, Sivan Ravid, and Chen Yanover. 'Framework for Identifying Drug Repurposing Candidates from Observational Healthcare Data'. JAMIA Open 3, no. 4 (31 December 2020): 536–44. https://doi.org/10.1093/jamiaopen/ooaa048.
- Austin, Peter C. 'An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies'. *Multivariate Behavioral Research* 46, no. 3 (May 2011): 399–424. https://doi.org/10.1080/00273171.2011.568786.