“The sole cause and root of almost every defect in the sciences is this: that whilst we falsely admire and extol the powers of the human mind, we do not search for its real helps.”

— Novum Organum: Aphorisms [Book One], 1620, Sir Francis Bacon
• **Public-Private Research Partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:**

  – Conducting methodological research to empirically evaluate the performance of various analytical methods on their ability to identify true associations and avoid false findings

  – Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum

  – Establishing a shared resource so that the broader research community can collaboratively advance the science
now called…

http://www.ohdsi.org
“Evidence-Based Medicine” as against ??
“In this dependence on the limited, idiosyncratic capacities of individuals, medical practice lags centuries behind the domains of science and commerce.”

- Lawrence L. Weed
Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort

Jane Green, clinical epidemiologist,1 Gabriela Czanner, statistician,1 Gillian Reeves, statistical epidemiologist,1 Joanna Watson, epidemiologist,1 Lesley Wise, manager, Pharmacoepidemiology Research and Intelligence Unit,2 Valerie Beral, professor of cancer epidemiology1

Conclusions The risk of oesophageal cancer increased with 10 or more prescriptions for oral bisphosphonates and with prescriptions over about a five year period.
BMJ study design choices

- Data source: General Practice Research Database
- Study design: Nested case-control
- Inclusion criteria: Age > 40
- Case: cancer diagnosis between 1995-2005 with 12-months of follow-up pre-diagnosis
- 5 controls per case
- Matched on age at index date, sex, practice, observation period prior to index
- Exposure definition: >=1 prescription during observation period
- “RR” estimated with conditional logistic regression
- Covariates: smoking, alcohol, BMI before outcome index date
- Sensitivity analyses:
  - exposure = 2+ prescriptions
  - covariates not missing
  - time-at-risk = >1 yr post-exposure
- Subgroup analyses:
  - Short vs. long exposure duration
  - Age, Sex, smoking, alcohol, BMI
  - Osteoporosis or osteopenia
  - Fracture pre-exposure
  - Prior diagnosis of Upper GI dx pre-exposure
  - NSAID, corticosteroid, H2blocker, PPI utilization pre-exposure
• In the design of observational studies we also rely heavily on “clinical judgment”

• Even worse, we do so with very limited feedback

• operating characteristics?

• Like early days of lab testing – “trust me, I measured it myself”
Do these choices matter?
Range of estimates across high-dimensional propensity score inception cohort (HDPS) parameter settings

Parameter settings explored in OMOP:
- **Washout period (1):** 180d
- **Surveillance window (3):** 30 days from exposure start; exposure + 30d; all time from exposure start
- **Covariate eligibility window (3):** 30 days prior to exposure, 180, all-time pre-exposure
- **# of confounders (2):** 100, 500 covariates used to estimate propensity score
- **Propensity strata (2):** 5, 20 strata
- **Analysis strategy (3):** Mantel-Haenszel stratification (MH), propensity score adjusted (PS), propensity strata adjusted (PS2)
- **Comparator cohort (2):** drugs with same indication, not in same class; most prevalent drug with same indication, not in same class
Range of estimates across univariate self-controlled case series (USCCS) parameter settings

USCCS Parameter settings explored in OMOP:
**Condition type (2):** first occurrence or all occurrences of outcome

**Defining exposure time-at-risk:**
**Days from exposure start (2):** should we include the drug start index date in the period at risk?
**Surveillance window (4):**
30 d from exposure start
Duration of exposure (drug era start through drug era end)
Duration of exposure + 30 d
Duration of exposure + 60 d

**Precision of Normal prior (4):** 0.5, 0.8, 1, 2

For Bisphosphonates-GI Ulcer hospitalization, USCCS using incident events, excluding the first day of exposure, and using large prior of 2:
- When surveillance window = length of exposure, no association is observed
- Adding 30d of time-at-risk to the end of exposure increased to a significant RR=1.14
Fix everything *except* the database...
Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer

Chris R. Cardwell, PhD
Christian C. Abnet, PhD
Marie M. Cantwell, PhD
Liam J. Murray, MD

Context  Use of oral bisphosphonates has increased dramatically in the United States and elsewhere. Esophagitis is a known adverse effect of bisphosphonate use, and recent reports suggest a link between bisphosphonate use and esophageal cancer, but this has not been robustly investigated.

Objective  To investigate the association between bisphosphonate use and esophageal cancer.

Conclusion  The use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer.
Pioglitazone and bladder cancer: a propensity score matched cohort study

Li Wei, Thomas M. MacDonald & Isla S. Mackenzie

Medicines Monitoring Unit (MEMO), Division of Medicine and Therapeutics, Ninewells Hospital and Medical School, Dundee, UK

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT
· Pioglitazone is mainly used in combination

CONCLUSION
The results suggest that pioglitazone may not be significantly associated with an increased risk of bladder cancer in patients with type 2 diabetes.
RESEARCH

The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case-control study

Laurent Azoulay assistant professor, Hui Yin statistician, Kristian B Filion assistant professor, Jonathan Assayag graduate student, Agnieszka Majdan endocrinologist, Michael N Pollak oncologist and professor, Samy Suissa professor

Conclusion The use of pioglitazone is associated with an increased risk of incident bladder cancer among people with type 2 diabetes.
The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case-control study

Laurent Azouley, assistant professor 1, 2, Hui Yin, statistician, Kristian Belfil, assistant professor 1, 3, Jonathan Assayag, graduate student 1, Agnieszka Majdan, endocrinologist 4, Michael N Pollak, oncologist and professor 2, Samy Suissa, professor 5
Does this stuff work at all?
OMOP 2010/2011 Research Experiment

- Open-source
- Standards-based

10 data sources
Claims and EHRs
200M+ lives
OSIM

OMOP Methods Library
- Inception cohort
- Case control
- Logistic regression

Common Data Model

- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

Legend
Total
True positive’ benefit
True positive’ risk
Negative control’

• 10 data sources
• Claims and EHRs
• 200M+ lives
• OSIM

Drug

ACE inhibitors
Amphotericin B
Antibiotics: erythromycin, sulfonamides, tetracyclines
Antiepileptics: carbamazepine, phenytoin
Benzodiazepines
Beta blockers
Bisphosphonates: alendronate
Tricyclic antidepressants
Typical antipsychotics
Warfarin

Outcome
Angioedema
Aplastic Anemia
Acute Liver Injury
Bleeding
Hip Fracture
Hospitalization
Myocardial Infarction
Mortality after MI
Renal Failure
GI Ulcer Hospitalization

Positives: 9
Negatives: 44
### Ground truth for OMOP 2011/2012 experiments

<table>
<thead>
<tr>
<th>Event</th>
<th>Positive controls</th>
<th>Negative controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
<td>118</td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>36</td>
<td>66</td>
<td>102</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
<td>88</td>
</tr>
<tr>
<td>Upper Gastrointestinal Bleeding</td>
<td>24</td>
<td>67</td>
<td>91</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>165</strong></td>
<td><strong>234</strong></td>
<td><strong>399</strong></td>
</tr>
</tbody>
</table>

**Criteria for positive controls:**
- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as ‘causative agent’ in Tisdale et al, 2010: “Drug-Induced Diseases”
- Literature review identified no powered studies with refuting evidence of effect

**Criteria for negative controls:**
- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as ‘causative agent’ in Tisdale et al, 2010: “Drug-Induced Diseases”
- Literature review identified no powered studies with evidence of potential positive association

Drug examples:
- isoniazid
- fluticasone
- indomethacin
- clindamycin
- ibuprofen
- loratadine
- sertraline
- pioglitazone
Exploring isoniazid and acute liver injury

Adverse events associated with treatment of latent tuberculosis in the general population

Benjamin M. Smith MD, Kevin Schwartzman MD MPH, Gillian Bartlett PhD, Dick Menzies MD MSc

ABSTRACT

Background: Guidelines recommend treatment of latent tuberculosis in patients at increased risk for active tuberculosis. Studies investigating the association of therapy with serious adverse events have not included the entire treated population nor accounted for comorbidities or occurrence of similar events in the untreated general population. Our objective was to estimate the risk of adverse events requiring hospital admission that were associated with therapy for latent tuberculosis infection in the general population.

Methods: Using administrative health data from the province of Quebec, we created a historical cohort of all residents dispensed therapy for latent tuberculosis between 1998 and 2003. Each patient was matched on age, sex and postal region with two untreated residents. The observation period was 18 months (from 6 months before to 12 months after initiation of therapy). The primary outcome was hospital admission for therapy-associated adverse events.

Results: During the period of observation, therapy for latent tuberculosis was dispensed to 9145 residents, of whom 95% started isoniazid and 5% started rifampin. Pretreatment comorbid illness was significantly more common among patients receiving such therapy compared with the matched untreated cohort. Of all patients dispensed therapy, 45 (0.5%) were admitted to hospital for a hepatic event compared with 15 (0.1%) of the untreated patients. For people over age 65 years, the odds of hospital admission for a hepatic event among patients treated for latent tuberculosis infection was significantly greater than among matched untreated people after adjustment for comorbidities (odds ratio [OR] 6.4, 95% CI 2.2–18.3). Excluding patients with comorbid illness, there were two excess admissions to hospital for hepatic events per 100 patients initiating therapy compared with the rate among untreated people over 65 years (95% CI 0.1–3.87).

Interpretation: The risk of adverse events requiring hospital admission increased significantly among patients over 65 years receiving treatment for latent tuberculosis infection. The decision to treat latent tuberculosis infection in elderly patients should be made after careful consideration of risks and benefits.
Smith et al. 2011 study design and results

- Data source: Administrative claims from health insurance board of Quebec
- Study design: Cohort
- Exposure: all patients dispensed >=30d of therapy, 180d washout
- Unexposed cohort: 2 patients per exposed, matched by age, gender, and region, with no tuberculosis therapy
- Time-at-risk: Length of exposure + 60 days
- Events: Incident hospital admission for noninfectious or toxic hepatitis
- “Event ratio” estimated with conditional logistic regression
- Covariates: prior hospitalization, Charlson score, comorbidities

<table>
<thead>
<tr>
<th>Outcome; age, yr</th>
<th>LTBI therapy cohort</th>
<th>Untreated cohort</th>
<th>Crude OR†</th>
<th>Adjusted OR‡</th>
<th>Adjusted OR§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>45/9145 (0.5)</td>
<td>15/18290 (0.1)</td>
<td>6.5 (3.8–11.1)</td>
<td>3.7 (2.0–6.9)</td>
<td>2.7 (1.3–5.6)</td>
</tr>
<tr>
<td>≤ 35</td>
<td>5/4523 (0.1)</td>
<td>1/9046 (0.0)</td>
<td>10.0 (1.2–85.6)</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>36–60</td>
<td>8/2533 (0.3)</td>
<td>7/5066 (0.1)</td>
<td>2.6 (1.0–6.9)</td>
<td>2.0 (0.6–6.9)</td>
<td>1.5 (0.4–5.6)</td>
</tr>
<tr>
<td>51–65</td>
<td>10/1232 (0.8)</td>
<td>4/2464 (0.2)</td>
<td>7.0 (2.3–21.3)</td>
<td>2.9 (0.7–13.0)</td>
<td>2.6 (0.4–16.0)</td>
</tr>
<tr>
<td>&gt; 65</td>
<td>22/857 (2.6)</td>
<td>3/1714 (0.2)</td>
<td>10.8 (4.2–28.0)</td>
<td>6.4 (2.2–18.3)</td>
<td>3.2 (0.9–11.7)</td>
</tr>
</tbody>
</table>
Revisiting the isoniazid – acute liver injury example

• Data source: MarketScan Medicare Beneficiaries (MDCR)
• Study design: Cohort
• Exposure: all patients dispensed new use of isoniazid, 180d washout
• Unexposed cohort: Patient with indicated diagnosis (e.g. pulmonary tuberculosis) but no exposure to isoniazid; negative control drug referents
• Time-at-risk: Length of exposure + 30 days, censored at incident events
• Covariates: age, sex, index year, Charlson score, number of prior visits, all prior medications, all comorbidities, all priority procedures
• “Odds ratio” estimated through propensity score stratification (20 strata)

What if this study design were applied consistently across all the positive and negative controls?
• ROC plots sensitivity vs. false positive rate
• Rank-order all pairs by RR from largest to smallest
• Calculate sensitivity and specificity at all possible RR thresholds

• Area under ROC curve (AUC) provides probability that method will score a randomly chosen true positive drug-outcome pair higher than a random unrelated drug-outcome pair
  • AUC=1 is perfect predictive model
  • AUC=0.50 is random guessing (diagonal line)
  • Cohort method on MDCR: AUC = 0.64

Isoniazid (RR=4.04):
  Sensitivity = 4%
  Specificity = 98%
Setting thresholds from an ROC curve

- Setting thresholds from an ROC curve

- If target sensitivity = 50%:
  - RR Threshold = 1.25
  - Specificity = 69%

- If threshold set to RR=2:
  - Sensitivity = 26%
  - Specificity = 90%

- If target specificity = 95%:
  - RR Threshold = 2.87
  - Sensitivity = 10%

- Cohort method on MDCR: AUC = 0.64
- AUC suggests that this method is modestly predictive, on the low end of diagnostic tests used in clinical practice, but at any given threshold there is a high false positive rate and/or false negative rate
- Question: what strategies can be applied to do even better?
Strategies to improve predictive accuracy

- Stratify results by outcome
- Tailor analysis to outcome
- Restrict to sufficient sample size
- Optimize analysis to the data source
Performance after applying these strategies

- Restricting to drugs with sufficient sample further increased AUC for all outcomes, but the degree of change varied by outcome
- Increased prediction comes as tradeoff with fewer drugs under surveillance
- Self-controlled cohort design continue to be optimal design, but specific settings changed in all outcomes
• All self-controlled designs (OS, ICTPD, SCCS) are consistently at or near the top of performance across all outcomes and sources
• Cohort and case-control designs have comparable performance, consistently lower than all self-controlled designs
• Substantial variability in performance across the optimal settings of each method
Optimal methods (AUC) by outcome and data source

<table>
<thead>
<tr>
<th>Data source</th>
<th>Acute kidney injury</th>
<th>Acute liver injury</th>
<th>Acute myocardial infarction</th>
<th>GI bleed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCR</td>
<td>OS: 401002 (0.92)</td>
<td>OS: 401002 (0.76)</td>
<td>OS: 407002 (0.84)</td>
<td>OS: 402002 (0.86)</td>
</tr>
<tr>
<td>CCAE</td>
<td>OS: 404002 (0.89)</td>
<td>OS: 403002 (0.79)</td>
<td>OS: 408013 (0.85)</td>
<td>SCCS: 1931010 (0.82)</td>
</tr>
<tr>
<td>MDCD</td>
<td>OS: 408013 (0.82)</td>
<td>OS: 409013 (0.77)</td>
<td>OS: 407004 (0.80)</td>
<td>OS: 401004 (0.87)</td>
</tr>
<tr>
<td>MSLR</td>
<td>SCCS: 1939009 (1.00)</td>
<td>OS: 406002 (0.84)</td>
<td>OS: 403002 (0.80)</td>
<td>OS: 403002 (0.83)</td>
</tr>
<tr>
<td>GE</td>
<td>SCCS: 1949010 (0.94)</td>
<td>OS: 409002 (0.77)</td>
<td>ICTPD: 3016001 (0.89)</td>
<td>ICTPD: 3034001 (0.89)</td>
</tr>
</tbody>
</table>

- Self-controlled designs are optimal across all outcomes and all sources, but the specific settings are different in each scenario
- AUC > 0.80 in all sources for acute kidney injury, acute MI, and GI bleed
- Acute liver injury has consistently lower predictive accuracy
- No evidence that any data source is consistently better or worse than others
Good performance?

• ...it all depends on your tolerance of false positives and false negatives...
• ...but we’ve created a tool to let you decide

http://elmo.omop.org
Takeaways from insights about risk identification

• Performance of different methods
  – Self-controlled designs appear to consistently perform well

• Evaluating alternative HOI definitions
  – Broader definitions have better coverage and comparable performance to more specific definitions

• Performance across different signal sizes
  – A risk identification system should confidently discriminate positive effects with RR>2 from negative controls

• Data source heterogeneity
  – Substantial variation in estimates across sources suggest replication has value but may result in conflicting results

• Method parameter sensitivity
  – Each method has parameters that are expected to be more sensitive than others, but all parameters can substantially shift some drug-outcome estimates
Revisiting clopidogrel & GI bleed (Opatrny, 2008)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Cases (n = 4028)</th>
<th>Controls (n = 40,171)</th>
<th>Crude rate ratio</th>
<th>Adjusted rate ratio*</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antidepressants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSRI</td>
<td>335 (8.3%)</td>
<td>1780 (4.4%)</td>
<td>1.97</td>
<td>1.33</td>
<td>1.09, 1.62</td>
</tr>
<tr>
<td>TCA</td>
<td>262 (6.5%)</td>
<td>1764 (4.4%)</td>
<td>1.52</td>
<td>1.04</td>
<td>0.83, 1.30</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>56 (1.4%)</td>
<td>229 (0.6%)</td>
<td>2.48</td>
<td>1.85</td>
<td>1.34, 2.55</td>
</tr>
<tr>
<td>Anticoagulant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warfarin</td>
<td>281 (7.0%)</td>
<td>1130 (2.8%)</td>
<td>2.64</td>
<td>2.17</td>
<td>1.82, 2.59</td>
</tr>
<tr>
<td>Clopidogrel</td>
<td>160 (4.0%)</td>
<td>532 (1.3%)</td>
<td>3.16</td>
<td>2.07</td>
<td>1.66, 2.58</td>
</tr>
</tbody>
</table>

OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)
Relative risk: 1.86, 95% CI: 1.79 – 1.93
Standard error: 0.02, p-value: <.001
Null distribution

CC: 2000314, CCAE, GI Bleed
Null distribution

CC: 2000314, CCAE, GI Bleed

Density

Relative Risk (Log scale)
Null distribution

CC: 2000314, CCAE, GI Bleed

clopidogrel
Evaluating the null distribution?

- Current p-value calculation assumes that you have an unbiased estimator (which means confounding either doesn’t exist or has been fully corrected for)

- Traditionally, we reject the null hypothesis at $p<.05$ and we assume this threshold will incorrectly reject the null hypothesis 5% of time. Does this hold true in observational studies?

- We can test this using our negative controls
## Ground truth for OMOP 2011/2012 experiments

<table>
<thead>
<tr>
<th>Event</th>
<th>Positive controls</th>
<th>Negative controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Liver Injury</td>
<td>81</td>
<td>37</td>
<td>118</td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>36</td>
<td>66</td>
<td>102</td>
</tr>
<tr>
<td>Acute Renal Failure</td>
<td>24</td>
<td>64</td>
<td>88</td>
</tr>
<tr>
<td>Upper Gastrointestinal Bleeding</td>
<td>24</td>
<td>67</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>165</td>
<td>234</td>
<td>399</td>
</tr>
</tbody>
</table>

Criteria for negative controls:
- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as ‘causative agent’ in Tisdale et al, 2010: “Drug-Induced Diseases”
- Literature review identified no evidence of potential positive association
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed

clopidogrel
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed

55% of these negative controls have p < .05
(Expected: 5%)
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed
Negative controls & the null distribution

CC: 2000314, CCAE, GI Bleed
p-value calibration plot
CC: 2000314, CCAE, GI Bleed
p-value calibration plot

CC: 2000314, CCAE, GI Bleed
p-value calibration plot

CC: 2000314, CCAE, GI Bleed
**p-value calibration plot**

CC: 2000314, CCAE, GI Bleed

- **clopidogrel:**
  - **RR** $1.9 (1.8 – 1.9)
  - **p** $< .001$
  - **Calibrated p** $0.30$

- **p < .05** $55%$
  - **Calibrated p < .05** $6%$
This analysis failed to reject the empirical null

... but we know clopidogrel causes GI bleeding (it’s a positive control)
p-value calibration plot

Optimal method: SCCS:1931010, CCAE, GI Bleed

- $p < .05$: 33%
- Calibrated $p < .05$: 9%
- clopidogrel:
  - RR: $1.3 (1.2 – 1.3)$
  - $p < .001$
  - Calibrated $p$: .01
Estimating coverage probability

- What if a study design could be applied across a large sample of drug-outcome pairs for which we know the true effect?
- Coverage probability: the percentage of the test cases where the estimated confidence interval contains the true effect (LB 95 CI <= true effect <= UB 95 CI)
- Challenge: in real data, the ‘true effect size’ for negative controls can be assumed to be RR=1, but the RRs for positive controls are not known
- In simulated data (OSIM2), we can inject signals with known effect sizes (RR=1.25, 1.50, 2, 4, 10) across a sample of drug-outcome scenarios and estimate the coverage probability
Applying case-control design to negative controls in real data

45% of the CIs of negative controls contain 1 (Expected: 95%)
Applying case-control design in simulated data, RR=1.0

75% of the CIs of negative controls contain 1 (Expected: 95%)
Applying case-control design to positive controls in simulated data, RR=1.25

54% coverage (Expected: 95%)
Applying case-control design to positive controls in simulated data, RR=1.50

46% coverage (Expected: 95%)
Applying case-control design to positive controls in simulated data, RR=2.00

42% coverage (Expected: 95%)
Applying case-control design to positive controls in simulated data, RR=4.00

25% coverage (Expected: 95%)
Applying case-control design and calibrating estimates of positive controls in simulated data, RR=1.25

Original coverage probability = 54%

Calibrated coverage probability = 96%
Applying case-control design and calibrating estimates of positive controls in simulated data, RR=1.50

Original coverage probability = 46%

Calibrated coverage probability = 92%
Applying case-control design and calibrating estimates of positive controls in simulated data, RR=2.00

Original coverage probability = 42%

Calibrated coverage probability = 92%
Applying case-control design and calibrating estimates of positive controls in simulated data, RR=4.00

Original coverage probability = 25%

Calibrated coverage probability = 100%
Coverage probability by effect size

Coverage probability by effect size

"True RR" – injected signal size

Coverage probability

"True RR" – injected signal size

Color by ESTIMATE_TYP:
- CALIBRATED
- ORIGINAL
Conclusions

• Using the OMOP approach, an empirical risk identification system can perform at AUC > 0.80
• Traditional p-values and confidence intervals require empirical calibration to account for bias in observational studies
• Advancing the science of observational research requires an empirical and reproducible approach to methodology and systematic application
• Predicting the effect of medical interventions is a causal inference problem