Dean’s Distinguished Lecture:
Drawing reproducible conclusions from observational clinical data

George Hripcsak, MD, MS
Biomedical Informatics, Columbia University
Medical Informatics Services, NewYork-Presbyterian
Disclosures

• No financial disclosures
Team

**PhD and Postdoc**
- Anna Ostropolets
- Linying Zhang
- Tara Anand
- Lauren Richter
- Ben Albert
- Tiffany Callahan

**OHDSI**
- Patrick Ryan
- Karthik Natarajan
- Soumitra Sengupta
- Marc Suchard
- Martijn Schuemie

**Causal Inference**
- Elias Bareinboim
- David Blei
- David Albers
- David Madigan
Observational research

- Subjects observed in their natural settings
  - Often using data collected for other purposes
  - Real-world evidence
- Versus experimental
  - Randomized clinical trials (RCTs)
- Administrative claims data
  - IBM Marketscan (10M’s)
- Electronic health record (EHR) data
  - Columbia clinical data warehouse (6M)
- Other sources
  - Census, social media, mobile sensors, imaging
US National EHR data, per year

- Healthcare $4,000,000,000,000 industry in US
  - can’t duplicate

<table>
<thead>
<tr>
<th>Count</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000,000,000</td>
<td>visit notes</td>
</tr>
<tr>
<td>35,000,000</td>
<td>admit notes, discharge sum.</td>
</tr>
<tr>
<td>46,000,000</td>
<td>procedure notes</td>
</tr>
<tr>
<td>3,000,000,000</td>
<td>prescriptions</td>
</tr>
<tr>
<td>1,000,000,000</td>
<td>laboratory tests</td>
</tr>
<tr>
<td>&gt;50,000,000,000</td>
<td>facts</td>
</tr>
</tbody>
</table>
Why large-scale analysis is needed in healthcare

Patrick Ryan
Patient-level predictions for personalized evidence requires big data

2 million patients seem excessive or unnecessary?

- Imagine a provider wants to compare her patient with other patients with the same gender (50%), in the same 10-year age group (10%), and with the same comorbidity of Type 2 diabetes (5%)

- Imagine the patient is concerned about the risk of ketoacidosis (0.5%) associated with two alternative treatments they are considering

- With 2 million patients, you’d only expect to observe 25 similar patients with the event, and would only be powered to observe a relative risk > 2.0

Aggregated data across a health system of 1,000 providers may contain 2,000,000 patients
But there is a catch

August 2010: “Among patients in the UK General Practice Research Database, the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer.”

Sept 2010: “In this large nested case-control study within a UK cohort [General Practice Research Database], we found a significantly increased risk of oesophageal cancer in people with previous prescriptions for oral bisphosphonates.”
Retracted COVID-19 papers


Eric J. Rubin, M.D., Ph.D.

On May 1, 2020, we published “Cardiovascular Dis and Mortality in Covid-19,” a study of the effect of preexisting angiotensin-converting enzyme (ACE) inhibitors and angiotensin (ARBs) on Covid-19. This retrospective study used data drawn from an i:

Comment
Expression of concern: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

The Lancet Editors

Show more 

Share Cite

https://doi.org/10.1016/S0140-6736(20)31290-3
Nonrandomized observational analyses of large electronic patient databases are being promoted as an alternative to randomized clinical trials as a source of “real-world evidence” about the efficacy and safety of new and existing treatments. For drugs or procedures that are already being used widely, such observational studies may involve exposure of large numbers of patients. Consequently, they have the potential to detect rare adverse effects that cannot plausibly be attributed to bias, generally because the relative risk is large (e.g., Reye’s syndrome associated with the use of aspirin, or rhabdomyolysis associated with the use of statin therapy). Nonrandomized clinical observation may also suffice to detect large beneficial effects when good outcomes would not otherwise be expected (e.g., control of diabetic ketoacidosis with insulin treatment, or the rapid shrinking of tumors with chemotherapy).

However, because of the potential biases inherent in observational studies, such studies cannot generally be trusted when — as is often the case — the effects of the treatment of interest are actually null or only moderate (i.e., less than a twofold difference in the incidence of the health outcome between using and not using the treatment). In those circumstances, large observational studies may yield misleading associations of a treatment with health outcomes that are statistically significant but noncausal, or that are mistakenly null when the treatment really does have clinically important effects. Instead, randomized, controlled trials of adequate size are generally required to ensure that any moderate benefits or moderate harms of a treatment are assessed reliably enough to guide patient care appropriately (Box 1).
False comparison

- NOT observational versus RCT
  - Don’t have the money to do the RCTs we need
  - Point-of-care randomization will help
- Observational versus expert opinion and instinct
  - Guidelines are mostly expert opinion

- We need to optimize our observational research
## Desired attributes for reliable evidence

<table>
<thead>
<tr>
<th>Desired attribute</th>
<th>Question</th>
<th>Researcher</th>
<th>Data</th>
<th>Analysis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Identical</td>
<td>Different</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Replicable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Similar</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Generalizable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Different</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Robust</td>
<td>Identical</td>
<td>Same or different</td>
<td>Same or different</td>
<td>Different</td>
<td>Similar</td>
</tr>
<tr>
<td>Calibrated</td>
<td>Similar (controls)</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Statistically consistent</td>
</tr>
</tbody>
</table>
Reliable evidence requires a new tool: the community

<table>
<thead>
<tr>
<th>Desired attribute</th>
<th>Question</th>
<th>Researcher</th>
<th>Data</th>
<th>Analysis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Identical</td>
<td>Different</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Replicable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Similar</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Generalizable</td>
<td>Identical</td>
<td>Different</td>
<td>Identical</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Robust</td>
<td>Identical</td>
<td>Same or different</td>
<td>Different</td>
<td>Identical</td>
<td>Similar</td>
</tr>
<tr>
<td>Calibrated</td>
<td>Similar (controls)</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Evidence sharing across community</td>
</tr>
</tbody>
</table>

- Community of researchers with important public health questions
- Data network using community standards
- Analyses sharing community open-source tools
- Application of community best practices for evaluation
- Evidence sharing across community
Observational Health Data Sciences and Informatics (OHDSI, as “Odyssey”)

Mission: To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care

A multi-stakeholder, interdisciplinary, international collaborative with a coordinating center at Columbia University

http://ohdsi.org
OHDSI Collaborators
- 2,367 collaborators
- 74 countries
- 21 time zones
- 6 continents
- 1 community

- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Currently records on about **800 million unique patients** in >300 databases
- 344 papers, specific influence on EMA and FDA for COVID-19
Open Science

Open science

Data + Analytics + Domain expertise

Generate evidence

Open source software

Enable users to do something

Standardized, transparent workflows

Database summary

Cohort definition

Cohort summary

Compare cohorts

Exposure-outcome summary

Effect estimation & calibration

Compare databases

Enable users to do something
How OHDSI Works

Source data warehouse, with identifiable patient-level data

ETL

Standardized, de-identified patient-level database (OMOP CDM v5)

Standardized large-scale analytics

Analysis results

OHDSI Data Partners

OHDSI Coordinating Center

Data network support

Analytics development and testing

Research and education

Summary statistics results repository

OHDSI.org
Current Approach: “One Study – One Script”

"What's the adherence to my drug in the data assets I own?"

Analytical method: Adherence to Drug

Application to data

Current solution:

Custom script for each study

- Not scalable
- Expensive
- Slow
- Prohibitive to non-expert routine use

Christian Reich
Solution: Standardized Data and Analytics

1. ATLAS, Remote Studies
   - Standard Cohorts
   - Standardized Analytics

2. OMOP CDM
   - Standardized Format
   - Standardized Coding
Extensive vocabularies
OHDSI’s standardized vocabularies

• 153 Vocabularies across 41 domains
  – MU3 standards: SNOMED, RxNorm, LOINC
  – Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS...

• >9 million concepts
  – >3.3 million standard concepts
  – >5.1 million source codes
  – >629,000 classification concepts

• >55 million concept relationships

• >84 million ancestral relationships

Publicly available for download at: http://athena.ohdsi.org/
Standardized conventions

Shared Conventions developed by the THEMIS Workgroup
Preparing your data

WhiteRabbit: profile your source data
RabbitInAHat: map your source structure to CDM tables and fields
ATHENA: standardized vocabularies for all CDM domains
Usagi: map your source codes to CDM vocabulary
CDM: DDL, index, constraints for Oracle, SQL Server, PostgreSQL
Vocabulary tables with loading scripts
ACHILLES: profile your CDM data; review data quality assessment; explore population-level summaries

OHDSI Forums:
Public discussions for OMOP CDM Implementers/developers

http://github.com/OHDSI
# Data Quality Dashboard

**IBM® MARKETSCAN® MULTI-STATE MEDICAID DATABASE**

DataQualityDashboard Version: 1.0.0

Results generated at 2020-08-24 15:44:34 in 3 hours

<table>
<thead>
<tr>
<th>STATUS</th>
<th>TABLE</th>
<th>CATEGORY</th>
<th>SUBCATEGORY</th>
<th>LEVEL</th>
<th>NOTES</th>
<th>DESCRIPTION</th>
<th>% RECORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIL</td>
<td>PAYER_PLAN_PERIOD</td>
<td>Conformance</td>
<td>Relational</td>
<td>FIELD</td>
<td>None</td>
<td>The number and percent of records that have a value in the payer_plan_period_id field in the PAYER_PLAN_PERIOD table that does not exist in the PERSON table. (Threshold=0%).</td>
<td>100.00%</td>
</tr>
<tr>
<td>FAIL</td>
<td>PROVIDER</td>
<td>Conformance</td>
<td>None</td>
<td>FIELD</td>
<td>None</td>
<td>The number and percent of records that do not have a standard, valid concept in the gender_concept_id field in the PROVIDER table. (Threshold=0%).</td>
<td>100.00%</td>
</tr>
<tr>
<td>PASS</td>
<td>PERSON</td>
<td>Completeness</td>
<td>None</td>
<td>FIELD</td>
<td>None</td>
<td>The number and percent of records with a NULL value in the birth_datetime of the PERSON. (Threshold=100%).</td>
<td>100.00%</td>
</tr>
<tr>
<td>PASS</td>
<td>PERSON</td>
<td>Completeness</td>
<td>None</td>
<td>FIELD</td>
<td>None</td>
<td>The number and percent of records with a NULL value in the provider_id of the PERSON. (Threshold=100%).</td>
<td>100.00%</td>
</tr>
<tr>
<td>PASS</td>
<td>PERSON</td>
<td>Completeness</td>
<td>None</td>
<td>FIELD</td>
<td>None</td>
<td>The number and percent of records with a NULL value in the care_site_id of the PERSON. (Threshold=100%).</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Showing 1 to 5 of 3,124 entries
ATLAS: Ontology support

- What terms do I need to create a cohort
- Tied to the database: what terms are used
  - Especially important for someone else’s database
ATLAS: Cohort building

- Optimized for observational research
  - Time series: who and *when* (vs classification)
  - Observation period, event timing
  - Assume a complex definition – Linearized AND-OR group
ATLAS: Analysis (observational)

- Approach: log regression, Poisson regression, survival
- Confounder: regularized-regression propensity score
- Residual confounding: calibration
- Diagnostics
ATLAS: Visualization

- Tables
- Graphs
Evidence OHDSI seeks to generate from observational data

• **Clinical characterization - tally**
  – Natural history: Who has diabetes, and who takes metformin?
  – Quality improvement: What proportion of patients with diabetes experience complications?

• **Population-level estimation - cause**
  – Safety surveillance: Does metformin cause lactic acidosis?
  – Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?

• **Patient-level prediction - predict**
  – Precision medicine: Given everything you know about me, if I take metformin, what is the chance I will get lactic acidosis?
  – Disease interception: Given everything you know about me, what is the chance I will develop diabetes?
Phenotyping
Meaning

- PERRLA

Pupils equal, round, reactive to light and accommodation
Data are mostly missing
  – Sampled when sick

Implicit information
  – Pertinent negatives by attending vs CC3
Noisy

- As low as 50% accuracy (Hogan JAMIA 1997)

- ... 36 year old man ... 27 year old woman ...
Complex

- Which is the right time?
  - When specimen drawn
  - When specimen received
  - When test performed
  - When result updated
  - When result received by the patient
  - When patient told clinician
  - When clinician wrote the note
Health care process bias
Good news

• Doctors successfully infer patient state from records

• We need to mimic the doctor’s reasoning
  – Deconvolve the truth
**EHR-derived phenotype**

- Clinically relevant feature derived from EHR
  - Patient has (a diagnosis of) type II diabetes
  - Recent rash and fever
  - Drug-induced liver injury
- Then use phenotype in correlation studies, ...
  - Which treatments associated with best outcomes
**EHR-derived phenotype**

- Want to know if patient has type 2 diabetes
  - Don’t just look up the disease in the record
  - Yes, look for diagnosis codes
  - Diabetes medications
  - Glucose suggestive of diabetes
  - Special diabetes tests
  - Diabetes complications
  - Mentions in notes
  - Exclusions like type 1 diabetes

- Can take months to define and test
OHDSI phenotyping pipeline

Phenotype library, literature

Prior work review

PHOEBE
Creating comprehensive concept set representing clinical idea

ATLAS
Creating cohorts of patients that satisfy inclusion and exclusion criteria

Cohort diagnostics
Examining cohorts

PheValuator
Computing phenotype performance metrics

Phenotype library
Storing phenotypes
PHOEBE: get the right concepts
Exploit vast data network: rate of every code everywhere

Ostropolets AMIA 2021
Phevaluator: automate the evaluation
Proxy for manual chart review

Swerdel JBI 2019
OHDSI in Action: Characterization
Treatment Pathways

Global stakeholders
- Public
- Academics
- Industry
- Regulator

Evidence
- RCT, Obs

Conduits
- Social media
- Lay press
- Literature
- Guidelines
- Advertising
- Formulary
- Labels

Inputs
- Indication
- Feasibility
- Cost
- Preference

Local stakeholders
- Family
- Patient
- Clinician
- Consultant

Local stakeholders
- Evidence
- Conduits

Inputs
- Global stakeholders
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Description</th>
<th>Population, millions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUSOM</td>
<td>Ajou University School of Medicine</td>
<td>South Korea; inpatient hospital EHR</td>
<td>2</td>
</tr>
<tr>
<td>CCAE</td>
<td>MarketScan Commercial Claims and Encounters</td>
<td>US private-payer claims</td>
<td>119</td>
</tr>
<tr>
<td>CPRD</td>
<td>UK Clinical Practice Research Datalink</td>
<td>UK; EHR from general practice</td>
<td>11</td>
</tr>
<tr>
<td>CUMC</td>
<td>Columbia University Medical Center</td>
<td>US; inpatient EHR</td>
<td>4</td>
</tr>
<tr>
<td>GE</td>
<td>GE Centricity</td>
<td>US; outpatient EHR</td>
<td>33</td>
</tr>
<tr>
<td>INPC</td>
<td>Regenstrief Institute, Indiana Network for Patient Care</td>
<td>US; integrated health exchange</td>
<td>15</td>
</tr>
<tr>
<td>JMDC</td>
<td>Japan Medical Data Center</td>
<td>Japan; private-payer claims</td>
<td>3</td>
</tr>
<tr>
<td>MDCD</td>
<td>MarketScan Medicaid Multi-State</td>
<td>US; public-payer claims</td>
<td>17</td>
</tr>
<tr>
<td>MDCR</td>
<td>MarketScan Medicare Supplemental and Coordination of Benefits</td>
<td>US; private and public-payer claims</td>
<td>9</td>
</tr>
<tr>
<td>OPTUM</td>
<td>Optum ClinFormatics</td>
<td>US; private-payer claims</td>
<td>40</td>
</tr>
<tr>
<td>STRIDE</td>
<td>Stanford Translational Research Integrated Database Environment</td>
<td>US; inpatient EHR</td>
<td>2</td>
</tr>
<tr>
<td>HKU</td>
<td>Hong Kong University</td>
<td>Hong Kong; EHR</td>
<td>1</td>
</tr>
</tbody>
</table>
Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak\textsuperscript{a,b,c,1}, Patrick B. Ryan\textsuperscript{c,d}, Jon D. Duke\textsuperscript{c,e}, Nigam H. Shah\textsuperscript{c,f}, Rae Woong Park\textsuperscript{c,g}, Vojtech Huser\textsuperscript{c,h}, Marc A. Suchard\textsuperscript{c,i,j,k}, Martijn J. Schuemie\textsuperscript{c,d}, Frank J. DeFalco\textsuperscript{c,d}, Adler Perotte\textsuperscript{a,c}, Juan M. Banda\textsuperscript{c,f}, Christian G. Reich\textsuperscript{c,j}, Lisa M. Schilling\textsuperscript{c,m}, Michael E. Matheny\textsuperscript{c,n,o}, Daniella Meeker\textsuperscript{c,p,q}, Nicole Pratt\textsuperscript{c,r}, and David Madigan\textsuperscript{c,s}

\textsuperscript{a}Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; \textsuperscript{b}Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; \textsuperscript{c}Observational Health Data Sciences and Informatics, New York, NY 10032; \textsuperscript{d}Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; \textsuperscript{e}Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; \textsuperscript{f}Center for Biomedical Informatics Research, Stanford University, CA 94305; \textsuperscript{g}Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; \textsuperscript{h}Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; \textsuperscript{i}Department of Biometrics, University of California, Los Angeles, CA 90095; \textsuperscript{j}Department of Biostatistics, University of California, Los Angeles, CA 90095; \textsuperscript{k}Department of Human Genetics, University of California, Los Angeles, CA 90095; \textsuperscript{l}Real World Evidence Solutions, IMS Health, Burlington, MA 01809; \textsuperscript{m}Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; \textsuperscript{n}Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; \textsuperscript{o}Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; \textsuperscript{p}Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; \textsuperscript{q}Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; \textsuperscript{r}Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and \textsuperscript{s}Department of Statistics, Columbia University, New York, NY 10027

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

Observational research promises to complement experimental research by providing large, diverse populations that would be infeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also can contribute to the design of experiments and inform the generalizability of experimental research. Understanding the diversity of populations

Without sufficiently broad databases available in the first stage, randomized trials are designed without explicit knowledge of actual disease status and treatment practice. Literature reviews are restricted to the population choices of previous investigations, and pilot studies usually are limited in scope. By exploiting the ClinicalTrials.gov national trial registry (9) and electronic health
Treatment pathways for diabetes

T2DM: All databases

- First drug
- Second drug
- Only drug
Population-level heterogeneity across systems, and patient-level heterogeneity within systems

Type 2 Diabetes Mellitus

- CCAE
- CPRD
- JMDC

Hypertension

- CUMC
- INPC

Depression

- MDCD
- GE
- OPTUM

Medications:
- Metformin
- Glimepiride
- Sitagliptin
- Rosiglitazone
- Glyburide
- Insulin, Glargine, Human
- Exenatide
- Liraglutide
- Insulin, Aspart, Human
- Saxagliptin
- Hydrochlorothiazide
- Lisinopril
- Metoprolol
- Amlodipine
- Furosemide
- Losartan
- Atenolol
- Valsartan
- Carvedilol
- Triamterene
- Diltiazem
- Ramipril
- Benazepril
- Olmesartan
- Spironolactone
- Clonidine
- Citalopram
- Bupropion
- Sertraline
- Escitalopram
- Fluoxetine
- Trazodone
- Venlafaxine
- Duloxetine
- Paroxetine
- Amitriptyline
- Mirtazapine
- Desvenlafaxine
- Nortriptyline
- Doxepin
25% of HTN patients (10% of others) have a unique path despite 250M pop
Conclusions: Network research

• It is feasible to encode the world population in a single data model
  – Over 10% now

• Generating evidence is feasible

• Stakeholders willing to share results

• Able to accommodate vast differences in privacy and research regulation
Population-level estimation

- Causal inference, hypothesis testing
- Creating reliable evidence
- OHDSI: study it scientifically
  - Distribution of study designs, parameters, databases, hypotheses
Standard error vs effect size

Statistically significant

Hazard ratio

p=0.05
Observational research results in literature

85% of exposure-outcome pairs have p < 0.05

29,982 estimates
11,758 papers

Schuemie Phil Trans A 2018
Observational research results in literature

Don’t know the denominator of negative studies.

29,982 estimates
11,758 papers
Observational research results in literature

29,982 estimates
11,758 papers
We’re not just guessing right
Observational research results in literature

• Individuals may produce good research studies

• In aggregate, the medical observational research system is a data dredging machine
Verified and open

**VERIFIED**

- Employ only previously validated methods
- Advanced, systematic methods to control bias
- Extensive diagnostics and large-scale controls
- Test many hypotheses to assess operating characteristics
- Study many databases, locations, practice types

**OPEN**

- Fully pre-specified public protocol
- All software open-source with public parameters
- All diagnostics made public with results initially blinded
- All results made publicly available
- Results paired with detailed attestation and characterization of populations studied

**Hypothesis**

**Study design**

**Raw analytic output**

**Validated result**

**Generalize to many populations**

Many hypotheses, one population

One hypothesis
10 LEGEND Principles (Large-scale Evidence Generation and Evaluation across a Network of Databases)

• LEGEND will generate evidence at a large scale
• Dissemination of the evidence will not depend on the estimated effects
• LEGEND will generate evidence using a prespecified analysis design
• LEGEND will generate evidence by consistently applying a systematic process across all research questions
  – No thumb on the scale
• LEGEND will generate evidence using best practices
• LEGEND will include empirical evaluation through the use of control questions
• LEGEND will generate evidence using open-source software that is freely available to all
• LEGEND will not be used to evaluate new methods
• LEGEND will generate evidence across a network of multiple databases
• LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network
1. Propensity score adjustment with large-scale covariate set: measured confounding (and some unmeasured?)
   - Take advantage of the huge databases and balance on tens of thousands of covariates, pulling in other variables (BP)
   - Mimic balance of randomization (imperfect)
   - Don’t rely on human expertise to select confounders: systematic
   - Diagnostics

Graham: “A standardized mean difference of ≤0.1 indicates a negligible difference.”
Confounding

- Does butane gas cause lung cancer?
Propensity score to address confounding

- Propensity score = probability of belonging to the target cohort vs. the comparator cohort, given the baseline covariates
- Propensity score can be used as a ‘balancing score’: if the two cohorts have similar propensity score distribution, then the distribution of covariates should be the similar
- Balance the propensity -> balance the covariates
- Balance the covariates -> the comparisons are similar
  - Make a causal assertion: must be due to the treatment

Rubin Biometrika 1983
How to select the confounders

• **Manual selection -> poor agreement**
  - Chien 2015: age, month, gender, #visits, income urbanization, #drugs, specific drugs, Charlson, comorbidities (16), +HDPS variables
  - Hicks 2018: age, sex, year of cohort entry, body mass index, smoking status, alcohol related disorders (including alcoholism, alcoholic cirrhosis of the liver, alcoholic hepatitis, and hepatic failure), and history of lung diseases (including pneumonia, tuberculosis, and chronic obstructive pulmonary disease), duration of HTN Rx, statin use, #drugs
  - Ku 2018: age, sex, race, income status, baseline HF, baseline myocardial infarction, baseline peripheral artery disease, baseline stroke, baseline eGFR, baseline proteinuria, and time-dependent covariates including diabetes mellitus, obesity, systolic blood pressure, statin use, aspirin use, diuretic use, and concurrent use of other antihypertensive agents for the outcome of HF
  - Magid 2010: age, gender, days on thiazide prior to 2nd agent start, # of visits prior to thiazide, Mean Systolic BP, Mean Diastolic BP, Chronic Obstructive Pulmonary Disease, Hyperlipidemia, Cancer, Dementia, Chronic liver disease, Depression
  - Hasvold 2014: age, gender, elevated blood glucose, overweight and low socio-economic status are known risk factors for diabetes, High cholesterol and hypertension are additionally known risk factors for CVD

• **Empirical selection**
Large-scale propensity score (LSPS)

• A **systematic** approach to propensity adjustment
• Use a large set of covariates (10,000 < n < 100,000)
• But don’t want to balance *everything*
  – Mediators – pre-treatment
  – Simple colliders – pre-treatment
  – Instruments – diagnostics, domain knowledge
  – M-bias – correlation with underlying causes
• Fit a propensity model
  – LASSO (regularized regression) because #variables > #cases
• Match or stratify on propensity score
• Diagnostic: check that covariate balance is achieved on all observed variables
Diagnostic: Covariate balance

Plot 60,000 covariates; most are binary:

\[
\text{abs}(P_{\text{target group}} - P_{\text{comparator group}}) \quad \text{standard deviation}
\]

Graham: “A standardized mean difference of \(\leq 0.1\) indicates a negligible difference.”
Diagnostic: equipoise

- What is the overlap between the groups
- If too small, poor generalizability and stability
What about confounding that is not measured?

• Some confounders are not directly measured but may be correlated with the many variables used by LSPS

• Hypertension study
  – Baseline blood pressure is an important confounder
  – But not measured in most databases, except Optum EHR database

Achieves (near) balance of BP despite its absence

<table>
<thead>
<tr>
<th></th>
<th>Systolic BP</th>
<th>Diastolic BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td>0.200</td>
<td>0.168</td>
</tr>
<tr>
<td>LSPS 60K</td>
<td>0.126</td>
<td>0.094</td>
</tr>
<tr>
<td>LSPS 60K+BP</td>
<td>0.046</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Hripcsak JAMIA IM 2020
Adjusting for other variables that are not directly measured

- We seem to balance lots of stuff beyond what was put into LSPS
  - Medications balance conditions (Dx)
  - Conditions balance medications (mostly)
  - Non-CV balances cardiovascular
  - Demographics do **not** work

<table>
<thead>
<tr>
<th>Test</th>
<th>Total Covariates</th>
<th># unbal. before matching</th>
<th># unbal. after matching</th>
<th>Max diff before</th>
<th>Max diff after</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full covariates</td>
<td>72,203</td>
<td>166</td>
<td>0</td>
<td>0.3975</td>
<td>0.035</td>
</tr>
<tr>
<td>Demographics only</td>
<td>72,203</td>
<td>166</td>
<td>158</td>
<td>0.3975</td>
<td>0.375</td>
</tr>
<tr>
<td>No conditions</td>
<td>72,203</td>
<td>166</td>
<td>0</td>
<td>0.3975</td>
<td>0.068</td>
</tr>
<tr>
<td>No drugs</td>
<td>72,203</td>
<td>166</td>
<td>8</td>
<td>0.3975</td>
<td>0.108</td>
</tr>
<tr>
<td>No procedures</td>
<td>72,203</td>
<td>166</td>
<td>0</td>
<td>0.3975</td>
<td>0.059</td>
</tr>
<tr>
<td>No cardiovascular-related concepts</td>
<td>72,203</td>
<td>166</td>
<td>0</td>
<td>0.3975</td>
<td>0.074</td>
</tr>
</tbody>
</table>
LSPS vs. manual selection on the effect of a missing confounder

- Lisinopril vs hydrochlorothiazide
  - Confounder type 2 diabetes
LSPS

- Reduce bias if balance on many covariates instead of a few human-selected covariates (bias measured via negative controls)

- LSPS performs better than competing methods like high-dimensional propensity score (HDPS)
2. Confidence interval calibration using negative controls: unmeasured confounding

- Address residual confounding using hypotheses you know the answer for
- If too many are positive, then systematic error is operative
- Calibrate to keep the type 1 error at 0.05
- Diagnostics
Negative controls

• Negative control
  – exposure-outcome where relative risk is believed to be 1
  – example: ingrowing nail

• OHDSI employs 50-100 negative controls
  – systematic methods allow large scale
All negative controls - adjusted

When using the propensity score, 16% have $p < 0.05$
After calibration, 4% have $p < 0.05$ (was 16%)
Addressing reproducibility #3

3. **Multiple databases, locations, practice types**
   - Look for consistency among databases, practices
   - Combine via meta-analysis
   - Aids generalizability

(Recent grant review)
Addressing reproducibility #4

4. Publish all hypotheses, code, parameters, runs
   • Pre-specify protocol so cannot cheat
   • Publish all code so that others can run it
   • Publish masked results, check diagnostics, reveal results

(Sharing source code)
5. Carry out on aligned hypotheses at scale
• Operating characteristics of the analysis
• Large-scale diagnostics
OHDSI results in line with expectations

11% of exposure-outcome pairs have calibrated p < 0.05
Large-scale estimation

• How to use it
  – Decide what question you are asking, then correct for multiple hypotheses

• Not “data-dredging”!
  – Data-dredging is not about what you do but about what you *throw out*
Only 29 different drugs in 5 different classes to choose from!

Distinguished from 28 drugs in 12 other classes that are classified as potential secondary agents (including Beta Blockers)

OHDSI “LEGEND” Hypertension Study
Filling in the evidence gaps


A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines

WRITING COMMITTEE MEMBERS
Paul K. Whelton, MB, MD, MSc, FAHA, Chair; Robert M. Carey, MD, FAHA, Vice Chair; Wilbert S. Aronow, MD, FAACC, FAHA*; Donald E. Casey, Jr, MD, MPH, MBA, FAHA*; Karen J. Collins, MBA*; Cheryl Danon, RN, BSN, MPH, PhD, FAHA*; Sontra M. DePauw, MHS, PA-C, CLS, AACC*; Samuel Gidding, MD, FAHA*; Kenneth J. Jamerson, MD*; Daniel W. Jones, MD, FAHA*; Eric J. MacLaughlin, PharmD**; Paul Muntner, PhD, FAHA*; Bruce Ovsiyanik, MD, MSc, MAS, MBA, FAHA*; Sidney C. Smith, Jr, MD, MACC, FAHA*; Crystal C. Spencer, JDO*; Randall S. Stafford, MD, PhD*; Sandra J. Taylor, MD, FAHA*; Randall J. Thomas, MD, MS, FAAC, FAHA*; Kim A. Williams, Sr, MD, MACC, FAHA*; Jeff D. Williamson, MD, MHSc#; Jackson T. Wright, Jr, MD, PhD, FAHA##

ACC/AHA TASK FORCE MEMBERS
Glenn N. Levine, MB, FAHA, Chair; Patrick O’Gara, MB, FAHA, MACC, Chair-Elect; Jonathan L. Halperin, MD, FACC, FAHA, Immediate Past Chair; Sana M. Al-Khatib, MD, MHS, FACC, FAHA; Joshua A. Beckman, MD, MS, FAHA; Kim K. Bincher, MS, PharmD, AACC; Riley R. Borgan, MD, PhD, FAHA, FAHA##; Ralph G. Brown, MD, MPH, Macc, FAHA; Lee A. Fleisher, MD, FACC, FAHA; Federico Gentile, MD, FACC; Samuel Gidding, MD, FAHA##; Zachary D. Goldberger, MD, MS, FACC, FAHA; Mark A. Eliotky, MD, FACC, FAHA; John Ilomemthi, MD, PhD, FAHA; José A. Jorgler, MD, FACC, FAHA; Laura Munt, MD, MSc, FAHA; Susan J. Pressler, RN, FAHA##; Barbara Riegel, PhD, RN, FAHA; Damiana N. Wijeyesundera, MD, PhD

Whelton et al., Hypertension 2018
Evidence to support the guideline

• 40 randomized trials
• Most decisions are expert opinion
## Comparisons of hypertension treatments

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ingredients</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Single ingredient comparisons</td>
<td>58 * 57 = 3,306</td>
<td>1,296</td>
</tr>
<tr>
<td>Single drug classes</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Single class comparisons</td>
<td>15 * 14 = 210</td>
<td>156</td>
</tr>
<tr>
<td>Dual ingredients</td>
<td>58 * 57 / 2 = 1,653</td>
<td>58</td>
</tr>
<tr>
<td>Single vs duo drug comparisons</td>
<td>58 * 1,653 = 95,874</td>
<td>3,810</td>
</tr>
<tr>
<td>Dual classes</td>
<td>15 * 14 / 2 = 105</td>
<td>32</td>
</tr>
<tr>
<td>Single vs duo class comparisons</td>
<td>15 * 105 = 1,575</td>
<td>832</td>
</tr>
<tr>
<td>Duo vs duo drug comparisons</td>
<td>1,653 * 1,652 = 2,730,756</td>
<td>2,784</td>
</tr>
<tr>
<td>Duo vs duo class comparisons</td>
<td>105 * 104 = 10,920</td>
<td>992</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Total comparisons</td>
<td>2,843,250</td>
<td>10,278</td>
</tr>
<tr>
<td>Outcomes of interest</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>Target-comparator-outcomes</td>
<td>2,843,250 * 58 = 164,908,500</td>
<td>587,020</td>
</tr>
</tbody>
</table>
Observational study to compare two initial therapies

Treatment strategies:
- Monotherapy with ACE
- Monotherapy with ARB

Causal contrasts of interest:
- Intent-to-treat effect
- On-treatment effect

Outcomes:
- Efficacy:
  - Myocardial infarction
  - Stroke
  - Heart Failure
- Safety:
  - Known or potential adverse events, e.g.
    - Acute renal failure
    - Angioedema
    - Cough
    - Diarrhea
    - Fall
    - Gout
    - Headache
    - Hyperkalemia
    - Hyponatremia
    - Hypotension
    - Impotence
    - Syncope
    - Vertigo

Eligibility criteria:
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

Analysis plan:
- Time-to-first-event analysis
- Cox proportional hazards

PS adjustment

Index: Time zero

Follow-up time

Medical history lookback time
58 outcomes of interest

Abdominal pain  Abnormal weight gain  Abnormal weight loss  Acute myocardial infarction  Acute pancreatitis  Acute renal failure  All-cause mortality  Anaphylactoid reaction  Anemia  Angioedema  Anxiety  Bradycardia  Cardiac arrhythmia  Cardiovascular disease  Cardiovascular-related mortality  Chest pain or angina  Chronic kidney disease  Coronary heart disease  Cough  Decreased libido  Dementia  Depression  Diarrhea  Edema  End stage renal disease  Fall  Gastrointestinal bleeding  Gout  Headache  Heart failure  Hemorrhagic stroke  Hepatic failure  Hospitalization with heart failure  Hospitalization with preinfarction syndrome  Hyperkalemia  Hypokalemia  Hypomagnesemia  Hyponatremia  Hypotension  Impotence  Ischemic stroke  Kidney disease  Malignant neoplasm  Measured renal dysfunction  Nausea  Neutropenia or agranulocytosis  Rash  Rhabdomyolysis  Stroke  Sudden cardiac death  Syncope  Thrombocytopenia  Transient ischemic attack  Type 2 diabetes mellitus  Vasculitis  Venous thromboembolic events  Vertigo  Vomiting
Abnormal cervical smear
Abnormal pupil
Abrasion and/or friction burn of trunk without infection
Absence of breast
Absent kidney
Acid reflux
Acquired hallux valgus
Acquired keratoderma
Acquired trigger finger
Acute conjunctivitis
Amputated foot
Anal and rectal polyp
Burn of forearm
Calcaneal spur
Cannabis abuse
Cervical somatic dysfunction
Changes in skin texture
Chondromalacia of patella
Cocaine abuse
Colostomy present
Complication due to Crohn's disease
Contact dermatitis
Contusion of knee
Crohn's disease
Derangement of knee
Difficulty sleeping
Disproportion of reconstructed breast
Effects of hunger
Endometriosis
Epidermoid cyst
Feces contents abnormal
Foreign body in orifice
Ganglion cyst
Genetic predisposition
Hammer toe
Hereditary thrombophilia
Herpes zoster without complication
High risk sexual behavior
Homocystinuria
Human papilloma virus infection
Ileostomy present
Impacted cerumen
Impingement syndrome of shoulder region
Ingrowing nail
Injury of knee
Irregular periods
Kwashiorkor
Late effect of contusion
Late effect of motor vehicle accident
Leukorrhea
Macular drusen
Melena
Nicotine dependence
Noise effects on inner ear
Non-specific tuberculin test reaction
Non-toxic multinodular goiter
Onychomycosis due to dermatophyte
Opioid abuse
Passing flatus
Postviral fatigue syndrome
Presbyopia
Problem related to lifestyle
Psychalgia
Ptotic breast
Regular astigmatism
Senile hyperkeratosis
Somatic dysfunction of lumbar region
Splinter of face, without major open wound
Sprain of ankle
Strain of rotator cuff capsule
Tear film insufficiency
Tobacco dependence syndrome
Vaginitis and vulvovaginitis
Verruca vulgaris
Wrist joint pain
Wristdrop
Databases

- **US insurance databases**
  - IBM® MarketScan® CCAE
  - IBM® MarketScan® MDCD
  - IBM® MarketScan® MDCR
  - Optum® Clinformatics®

- **Japanese insurance database**
  - Japan Medical Data Center

- **Korean national insurance database**
  - NHIS-NSC

- **US EHR databases**
  - Columbia University Irving Medical Center
  - Optum® PANTHER®

- **German EHR database**
  - QuintilesIMS Disease Analyzer (DA) Germany
Efficacy outcome: *myocardial infarction*, heart failure, stroke

<table>
<thead>
<tr>
<th>RCTs</th>
<th>LEGEND</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEIs</td>
<td>ACEIs</td>
</tr>
<tr>
<td>ARBs</td>
<td>ARBs</td>
</tr>
<tr>
<td>cBBs</td>
<td>cBBs</td>
</tr>
<tr>
<td>dCCBs</td>
<td>dCCBs</td>
</tr>
<tr>
<td>TZDs</td>
<td>TZDs</td>
</tr>
</tbody>
</table>

Data source: meta-analysis, $\sim 1-2$M total patients per study

- Beta blockers underperform alternatives
- Unexpected: TZDs > ACEIs. Reliable?

*Lancet* 2019
Prescriptions are not written at the class-level; must choose an individual drug for the patient.

- $1^{\text{st}}$-line > $2^{\text{nd}}$-line
- Some within-class differences failed diagnostics, e.g. captopril

Composite (MI, HF, stroke) outcome in meta-analysis
Chlorthalidone vs hydrochlorothiazide: worse safety without real world effectiveness

Risk estimates and meta-analysis across LEGEND databases:

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Data source</th>
<th>HR</th>
<th>LB</th>
<th>UB</th>
<th>P</th>
<th>Cal.HR</th>
<th>Cal.LB</th>
<th>Cal.UB</th>
<th>Cal.P</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS stratification, on-treatment</td>
<td>CCAE</td>
<td>0.65</td>
<td>0.33</td>
<td>1.14</td>
<td>0.17</td>
<td>0.66</td>
<td>0.37</td>
<td>1.19</td>
<td>0.18</td>
</tr>
<tr>
<td>PS stratification, on-treatment</td>
<td>Meta-analysis</td>
<td>0.79</td>
<td>0.54</td>
<td>1.16</td>
<td>0.24</td>
<td>0.81</td>
<td>0.56</td>
<td>1.17</td>
<td>0.30</td>
</tr>
<tr>
<td>PS stratification, on-treatment</td>
<td>Optum</td>
<td>0.90</td>
<td>0.52</td>
<td>1.44</td>
<td>0.67</td>
<td>0.93</td>
<td>0.57</td>
<td>1.53</td>
<td>0.82</td>
</tr>
<tr>
<td>PS stratification, on-treatment</td>
<td>Panther</td>
<td>0.98</td>
<td>0.05</td>
<td>5.06</td>
<td>0.99</td>
<td>0.91</td>
<td>0.26</td>
<td>3.42</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Table 1a.** Number of subjects, follow-up time (in years), number of outcome events, and event incidence rate (IR) per 1,000 patient years (PY) in the target (Chlorthalidone) and comparator (Hydrochlorothiazide) group after stratification, as well as the minimum detectable relative risk (MDRR). Note that the IR does not account for any stratification.
CTD vs. HCTZ: safety profile

- Safety favors HCTZ – electrolyte imbalance
- CTD is more potent, longer half-life
HCTZ vs chlorthalidone

- **Physiology**
  - Chlorthalidone is more potent and longer lasting
- **Indirect (network) meta-analysis favors chlorthalidone**
  - Combine RCT results
  - Bias: heterogeneity of treatment effect + different populations
  - Also: differential RCT design and execution
- **Recent observational research favors HCTZ**
- **VA Diuretic Comparison Project**
  - RCT with completion 2023
  - Different question: of patients tolerating HCTZ, should they switch to chlorthalidone
- **Response**
  - 50% failure off diuretic (chlorthalidone is faster) in 2 months
  - Time at risk too short (but 25% are long term); do have sufficient power
  - Anecdotes
- **Chlorthalidone is a potent drug**

Choice of drug therapy in primary (essential) hypertension – UpToDate

By contrast, other observational studies suggest that chlorthalidone and hydrochlorothiazide lead to similar rates of cardiovascular events but that chlorthalidone increases the risk of adverse metabolic effects [35,36].

Based upon the above observations, we and other experts suggest that thiazide-like diuretics (such as chlorthalidone...
ACEi versus ARB

On the other hand, once the field is confident in a result...

Hypertension

ORIGINAL ARTICLE

Comparative First-Line Effectiveness and Safety of ACE (Angiotensin-Converting Enzyme) Inhibitors and Angiotensin Receptor Blockers

A Multinational Cohort Study

Rui Jun Chen, Marc A. Suchard, Harlan M. Krumholz, Martin J. Schuenke, Steven Shaw, Jon Duke, Nicole Pratt, Christian G. Reich, David Madigan, Seng Chan You, Patrick B. Ryan, George Hripcsak

ABSTRACT: ACE (angiotensin-converting enzyme) inhibitors and angiotensin receptor blockers (ARBs) are equally guideline-recommended first-line treatments for hypertension, yet few head-to-head studies exist. We compared the real-world effectiveness and safety of ACE inhibitors versus ARBs in the first-line treatment of hypertension. A large, prospective, non-randomized, real-world cohort study was conducted with data from 199,011 patients with hypertension from 12 centers in the USA, Canada, and Europe. ACEi and ARB were equally effective in reducing mean arterial pressure in patients starting treatment. ACEi were more expensive and associated with significantly higher risk of cardiovascular events, death, stroke, and fresh myocardial infarction. ARB were associated with significantly lower risk of acute respiratory infections compared to ACEi. The study’s authors write that clinical guidelines should be updated to include real-world evidence on the effectiveness and safety of ACE inhibitors versus ARBs and that randomized controlled trials of ACE inhibitors versus ARBs in the first-line treatment of hypertension are needed.

Circulation

Volume 145, Issue 6, 8 February 2022; Pages 413-415
https://doi.org/10.1161/CIRCULATIONAHA.121.057835

ON MY MIND

Why Are We Still Prescribing Angiotensin-Converting Enzyme Inhibitors?

Franz H. Messerli, MD, Chirag Bavishi, MD, MPH, and Sripal Bangalore, MD, MHA

The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it.

—Sir Francis Bacon, 1620
LEGEND vs RCTs: 28/30

Diagram showing hazard ratios for various conditions (Heart failure, Myocardial infarction, Stroke) across different treatment groups (ACE, ARB, BB, CCB). The diagram includes data from randomized clinical trials meta-analysis and LEGEND real-world evidence meta-analysis, with concordance indicated by reference lines and disagreement highlighted with statistically significant differences (p < 0.05).
OHDSI “LEGEND” Hypertension Study
Filling in the evidence gaps

Head-to-head HTN drug comparisons

- **Trials:** 40
- **N =** 102 – [1148] – 33K

- **Comparisons:** 10,278
- **N =** 3502 – [212K] – 1.9M
LEGEND Hypertension
OHDSI predictive algorithm validation framework

Clear specification of the prediction task:
- Target Population: patients at risk
- Outcome: medical event to predict
- Time-at-risk (TAR): interval to predict if outcome will occur

The patient-level prediction journey is more than just classification...

Join the PLP journey
PLP GitHub: github.com/OHDSI/PatientLevelPrediction

Design and Extraction
- Study design
  Case-control prone to misclassification and should be avoided; use cohort design

Model Development
- Learning across datasets
  Models can be learned across datasets while maintaining privacy

Model Evaluation
- Model usability
  Simple score-based models are easier to apply and can be benchmarked against large-scale models

- Visualizing performance
  A simple plot with the operating characteristics for all cut-offs informs model usefulness

- Network validation
  The OHDSI network enables large-scale external validation and improves our understanding of models

- Sample size
  Learning curves provide a way for model developers to determine whether they have sufficiently sized data

- Feature extraction
  Feature lookback can make an impact on model performance if it is too short (<180 days)
Evidence was needed around the use of hydroxychloroquine (HCQ) alone and in combination with azithromycin (AZ). We examined the use of these drugs in rheumatoid arthritis (RA) patients.

Findings:
- In history use in RA population, HCQ alone is generally safe but in combination with AZ it shows a doubling of risk of 30-day cardiovascular mortality.
Patients with cardiovascular diseases and hypertension treated with angiotensin converting enzyme inhibitors (ACEs) angiotensin-II receptor blockers (ARBs) may influence susceptibility to COVID-19 and worsen its severity.

As stated by Watson et al., in relation to one of the published studies, lack of transparency and uncertainties about research standards applied raise doubts about published results. Morales et al. supported the reproducibility of their study by publishing the study protocol in the EU PAS Register ahead of time, providing a start-to-finish executable code, facilitating the sharing and exploration of the complete result set with an interactive web application and asking clinicians and epidemiologists to perform a blinded evaluation of propensity score diagnostics for the treatment comparisons.
COVID-19 Vaccine Safety Methods Research

- AstraZeneca vaccine
  - March 11-15, 2021 – 13 European countries suspend use for fears of blood clots
    - Denmark, Norway, Iceland, Bulgaria, Ireland, Netherlands, Spain, Germany, Italy, France, ...
  - March 18, 2021 – EMA determines benefits outweigh the risks
    - Thromboembolic events “lower than that expected in the general population”
    - DIC and CVST above baseline but very rare
    - “The number of reported events exceeds those expected, and causality although not confirmed, cannot therefore be excluded. However, given the rarity of the events, and the difficulty of establishing baseline incidence since COVID-19 itself is resulting in hospitalisations with thromboembolic complications, the strength of any association is uncertain.”

- Partnered with FDA Center for Biologics Evaluation and Research (CBER)
  - Vaccine safety methods research, network and local studies
Standards enabling evidence for policy: COVID-19 treatment utilization patterns

Use of repurposed and adjuvant drugs in hospital patients with COVID-19: multinational network cohort study

Albert Prats-Uribe, Anthony G Sena, Lan Yin Hui Lai, Hegh Alghoul, Osaid Alser, Thamir M Alshammar, Paula Casajust, Dalia Dawoud, Asieh Golozar, Paras P Mehta, Mengchun Gong, Daniel R Morales, Martina Recalde, Elena Reol, Karishma Shah, Vignesh Subbian, David Vizcaya, Lin Zhang, Ying Jae Yeo, Cho, Kristine E Lynch, Michael E Matheny, Peter R Rijnekeek, George Hripcev, Jennifer CE Lane, Marc A Suchard, Talita Duarte-Salles, Kristin Kostka, Dani Nieto-Alhambra

ABSTRACT
OBJECTIVE
To investigate the use of repurposed and adjuvant drugs in patients admitted to hospital with COVID-19 across five continents.

DESIGN
Multinational network cohort study.

SETTING
Hospital electronic health records from the United States, Spain, China, and nationwide claims data from South Korea.

PARTICIPANTS
303,264 patients admitted to hospital with COVID-19 from January 2020 to December 2020.

MAIN OUTCOME MEASURES
Prescriptions or dispensations of any drug on or 30 days after the date of hospital admission for COVID-19.

RESULTS
Of the 303,264 patients included, 290,131 were from the US, 7,599 from South Korea, 5230 from Spain, and 304 from China. 3,455 drugs were identified. Common repurposed drugs were hydroxychloroquine, lopinavir/ritonavir, and umifenovir. Antidepressants, antiarrhythmics, and anticoagulants were frequently used.

CONCLUSIONS
The use of repurposed and adjuvant drugs varied across continents and may have contributed to the observed differences in outcomes. Further research is needed to better understand the comparative risk and benefit of these treatments in the management of COVID-19.

Fig 4 | Time trends in hydroxychloroquine use on days 0 to 30 after hospital admission in patients with a positive test result for or diagnosis of COVID-19 by month. CUIMC=Columbia University Irving Medical Center; HIRA=Health Insurance Review and Assessment; OMOP=Observational Medical Outcomes Partnership; Optum-EHR=Optum deidentified electronic health record dataset; STARR=Stanford medicine Research data Repository; VA=Veterans Affairs
Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study

Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study

Renin-angiotensin system blockers and susceptibility to COVID-19: an international, open science, cohort analysis

Summary
Background Angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin receptor blockers (ARBs) have been postulated to affect susceptibility to COVID-19. Observational studies so far have lacked rigorous ascertainment adjustment and internal-external confirmation with an increased susceptible

COVID-19

41 studies in the past year
OHDSI collaborations
US Food and Drug Administration CBER Biologics Effectiveness and Safety (BEST) OHDSI serves as the Convener

INTRODUCTION
The U.S. FDA Center for Biologics Evaluation and Research (CBER) regulates collection of whole blood and blood components utilized in transfusion.

OBJECTIVE
The aim of this study was to build a component of the infrastructure for a national hemovigilance system using EHR data sources to monitor transfusion-related AEs by incorporating the ISBT-128 coding system into the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) of the Observational Health Data Sciences and Informatics (OHDSI) consortium.

METHODS
The CBER BEST Initiative is a collaboration with IQVIA, OHDSI Consortium, Columbia University, Stanford University, Indiana University, Regenstrief Institute, Georgia Institute of Technology, and University of California Los Angeles. Within the BEST Initiative, we used three EHR databases that cover approximately 24 million patient records from geographically diverse areas of the U.S. We added a library of 14,548 ISBT-128 codes to the OMOP CDM. Each EHR data source requested access to its corresponding blood bank data and transformed its data into the OMOP CDM containing the newly added ISBT-128 codes. By querying the databases, we determined the type and frequency of ISBT-128 codes used in patient records from 2010-2017 within the blood banks of EHR data providers participating in the BEST Initiative.
US National Institutes of Health

All of Us Research Program

• 1,000,000 diverse participants
• Clinical data in OMOP CDM
• $100Ms

The future of health begins with you

The All of Us Research Program is a historic effort to gather data from one million or more people living in the United States to accelerate research and improve health. By taking into account individual differences in lifestyle, environment, and biology, researchers will uncover paths toward delivering precision medicine.

JOIN NOW
US National Institutes of Health
Electronic Medical Records and Genomics (eMERGE)
The N3C is a partnership among the NCATS-supported Clinical and Translational Science Awards (CTSA) Program hubs and the National Center for Data to Health (CD2H), with overall stewardship by NCATS. Collaborators will contribute and use COVID-19 clinical data to answer critical research questions to address the pandemic.

Building an Innovative Analytics Platform to Study COVID-19

The N3C is a new effort that aims to build a centralized national data resource that the research community can use to study COVID-19 and identify potential treatments as the pandemic continues to evolve.
The European Health Data and Evidence Network (EHDEN) 30M €
Innovative Medicines Initiative

**Mission**
Our mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardised to a common data model.

**Vision**
The European Health Data & Evidence Network (EHDEN) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care.

DARWIN EU – European Medicines drug surveillance initiative
Erasmus MC using OHDSI
National CDM Projects in Korea
OHDSI Data Network

70% of Tertiary Teaching Hospitals
Structural causal models
Bareinboim, Blei, Zhang, Anand

Mechanistic models
Albers, Richter, Albert

Ontologies
Callahan
OHDSI at CUIMC
iNYP -> Explore (Epic link)

- “De-identified” OMOP
  - refreshed 1-2 times a year

- Access to 15K users via clinical system
  - “Can I have this for my patient list”

- Use it to collaborate with clinical researchers
  - “Data consultation” ~10/yr
  - Improves our data
Columbia Data Consult Service

• Research project to study the effect of real-time evidence generation
• Put observational research into action
• 29 questions, 22 clinicians, 24 answers
  – Largely medicine, but due to recruitment
  – Mostly were about recurring issues
  – A fifth about a specific patient

Ostropolets JAMIA 2021
Observational Research Task Force

• Dawn Hershman, George Hripcsak, co-chairs
• Membership being confirmed
  – Clinical researchers, epidemiology, equity, informatics, privacy, health records

• Components
  – Research themes for emphasis
  – Faculty development and growth in these areas of research
  – Multi-PI collaborations
  – Education/training
  – Linkage to other initiatives and their task forces’ reports: e.g. core facilities (completed); clinical trials (completed); biostatistics (Under Dr. Kiros Berhane, ongoing).
Summary

• Current observational research is suspect
• Large-scale observational research appears to be possible and more reliable than the current approach
• 6-million-patient database is available for research today (and Marketscan with funding)

Funding
National Library of Medicine
R01 LM006910 (2000-2024)
T15 LM007079 (1994-2027)
NIDDK R01 DK090372
NCI contracts
NHGRI All of Us
FDA CBER BEST