An Open Benchmark for Causal Inference Using the MIMIC-III and Philips Datasets

Leo Anthony Celi, M.D. Ph.D. [1], Ken Jung, Ph.D. [2], Marzyeh Ghassemi, Ph.D. [3], Carlos Guzman[4], Uri Shalit, Ph.D. [4], David Sontag, Ph.D. [4]

[1]Laboratory for Computational Physiology, Massachusetts Institute of Technology; [2]Stanford Center for Biomedical Informatics, Stanford University; [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology; [4]Computer Science Department, Courant Institute for Mathematical Sciences, New York University

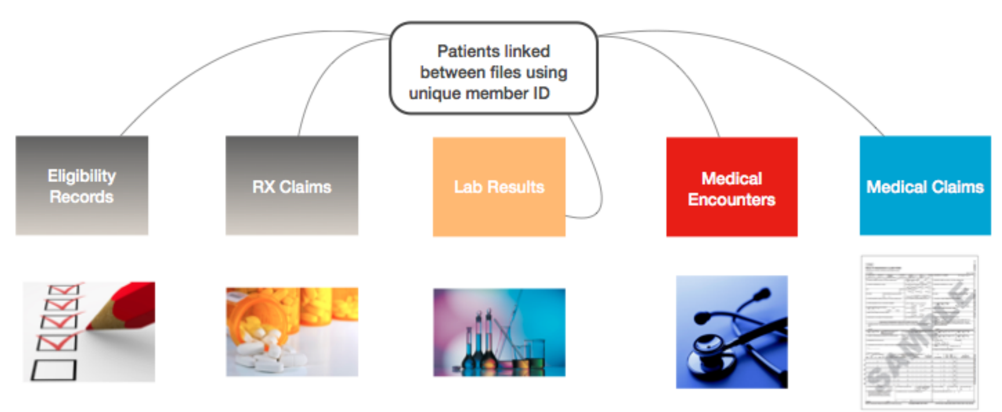
Contact: uas1@nyu.edu

Goal

Create an open, public benchmark for causal inference from observational hospital data

Background

- Large scale observational datasets: promise of inferring new causal relationships
- Massive and diverse datasets → require new statistical methods
- Challenge: how to compare and evaluate observational causal inference methods?
- Solution: identify RCTs conducted within observational datasets



Open hospital ICU datasets

- Two public de-identified datasets: MIMIC-III and Philips (released early 2017)
- More than 250,000 ICU admissions of ~200,000 adults over 11 years at dozens of hospitals across the USA
- Demographics, vital sign measurements made at the bedside (~1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, mortality (both in and out of hospital)

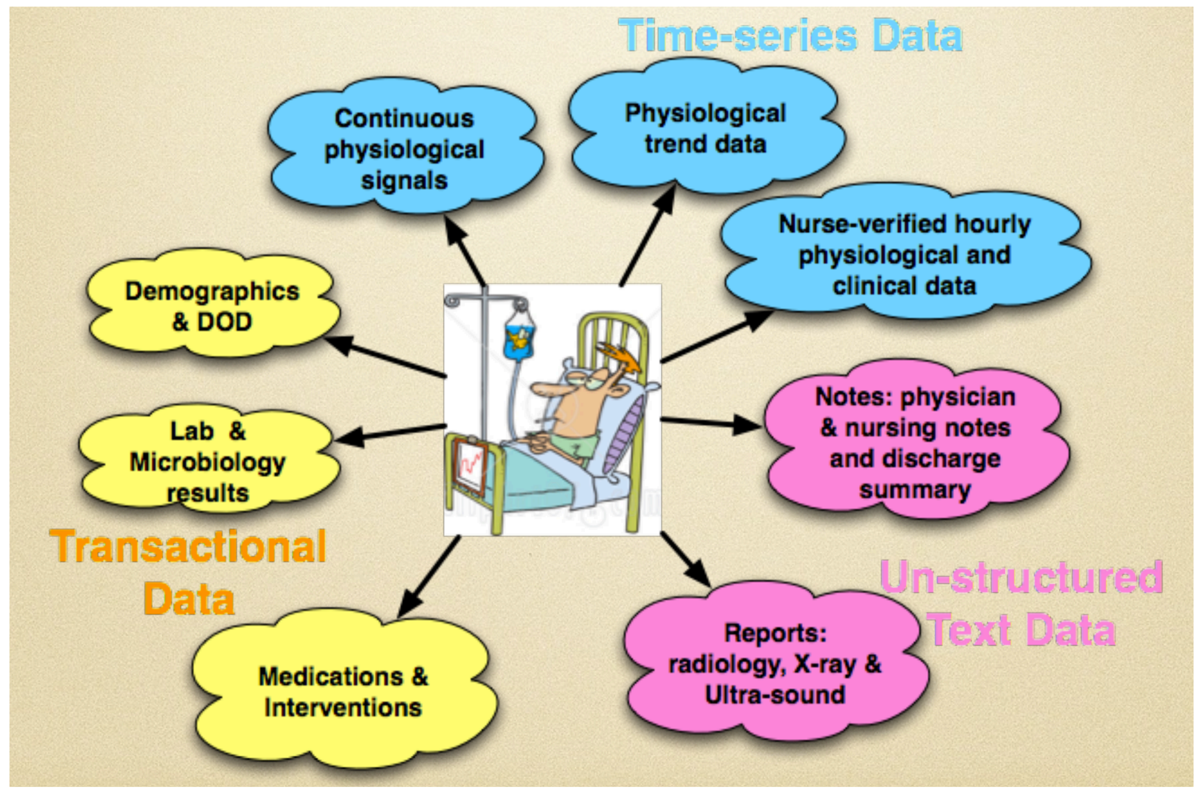


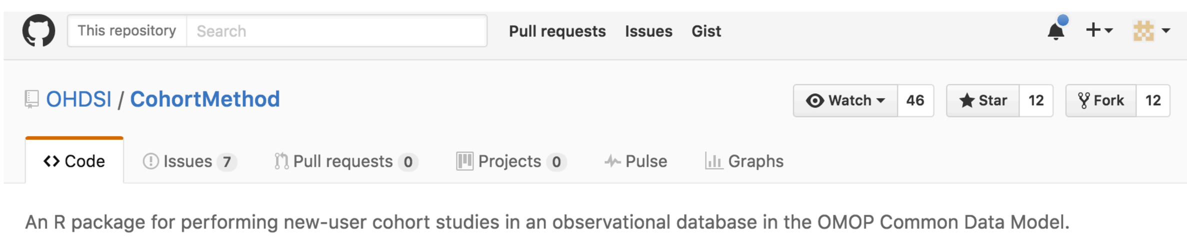
Image: Mengling Feng

Benchmarking Causal Inference for Observational Studies

	Real-world confounders	Real-world treatment assignment	Real-world outcomes	Compare study designs	Public
2016 Atlantic Causal Inference Competition	✓	✗	✗	✗	✓
Adverse drug reaction, Ryan et al. (2012)	✓	✓	✓	✓	✗
Proposed benchmark	✓	✓	✓	✗	✓

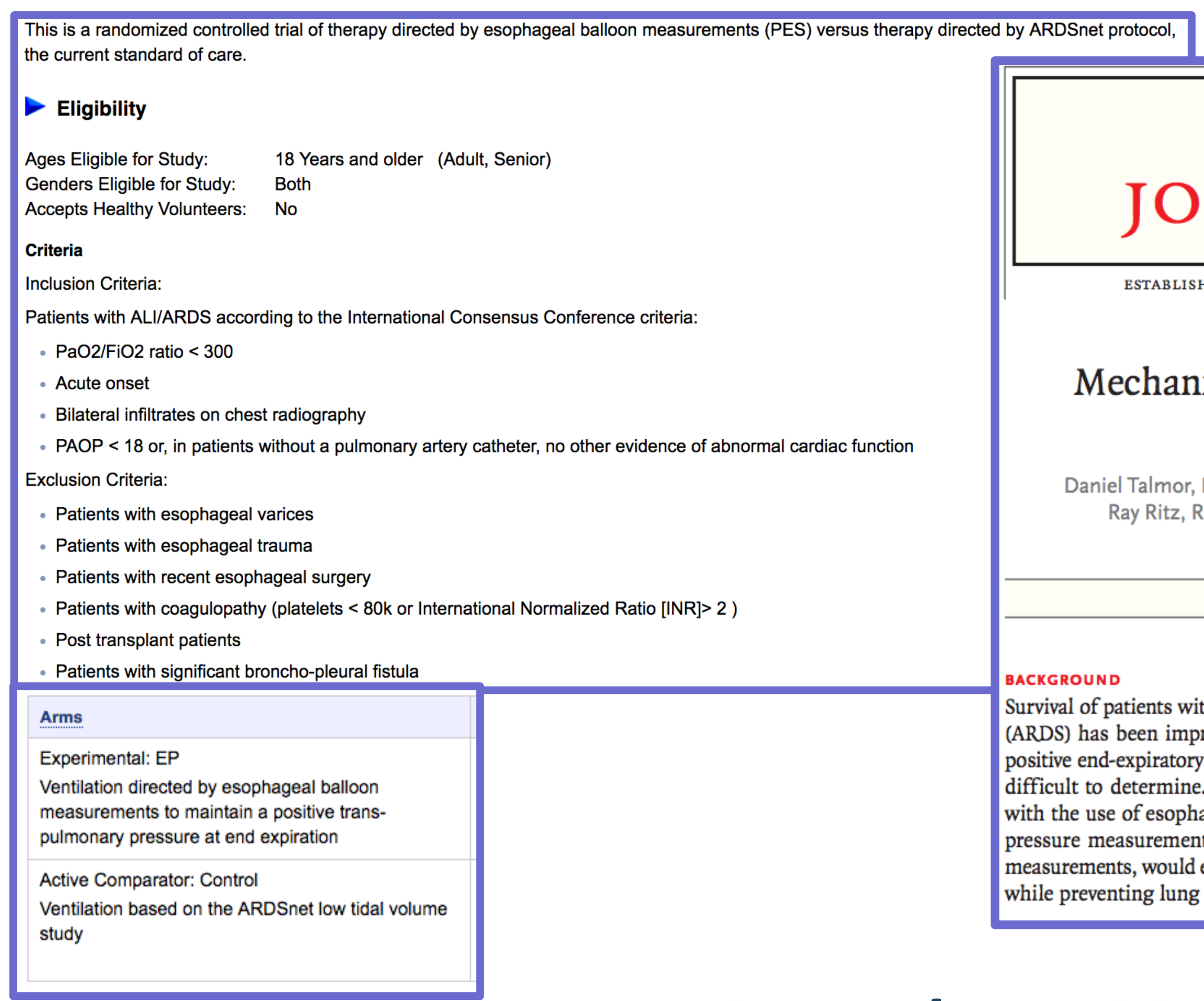
Methods

- Identify RCTs performed in ICUs similar to the MIMIC/Phillips ICUs
- Identify cohorts equivalent to the RCT cohorts within the public datasets using OMOP CDM
- Only small subset of eligible RCTs can plausibly be replicated
  - Treatment, outcome, inclusion/exclusion criteria, important background variables (confounders)
- For each identified RCT: treatment ($T = 0, T = 1$) and outcome Y
  Average Treatment Effect ATE $= \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$
- Identify potential confounders in observational cohort
- Create and publish easily accessible datasets for researchers outside of the medical informatics community, in two forms:
  - *(Confounder, Treatment, Outcome)*
  - *Cohort creation script*
- Baseline observational study
  - Based on OHDSI CohortMethod, propensity score and matching
  - Non-linear methods such as Bayesian Additive Regression Trees
- In 2017 run a public competition for inferring the RCT ATE from the published observational datasets



Candidate studies (preliminary)

clinicaltrials.gov



This is a randomized controlled trial of therapy directed by esophageal balloon measurements (PES) versus therapy directed by ARDSnet protocol, the current standard of care.

Eligibility

Ages Eligible for Study: 18 Years and older (Adult, Senior)
Genders Eligible for Study: Both
Accepts Healthy Volunteers: No

Criteria

Inclusion Criteria:

Patients with ALI/ARDS according to the International Consensus Conference criteria:

- PaO2/FiO2 ratio < 300
- Acute onset
- Bilateral infiltrates on chest radiography
- PAOP < 18 or, in patients without a pulmonary artery catheter, no other evidence of abnormal cardiac function

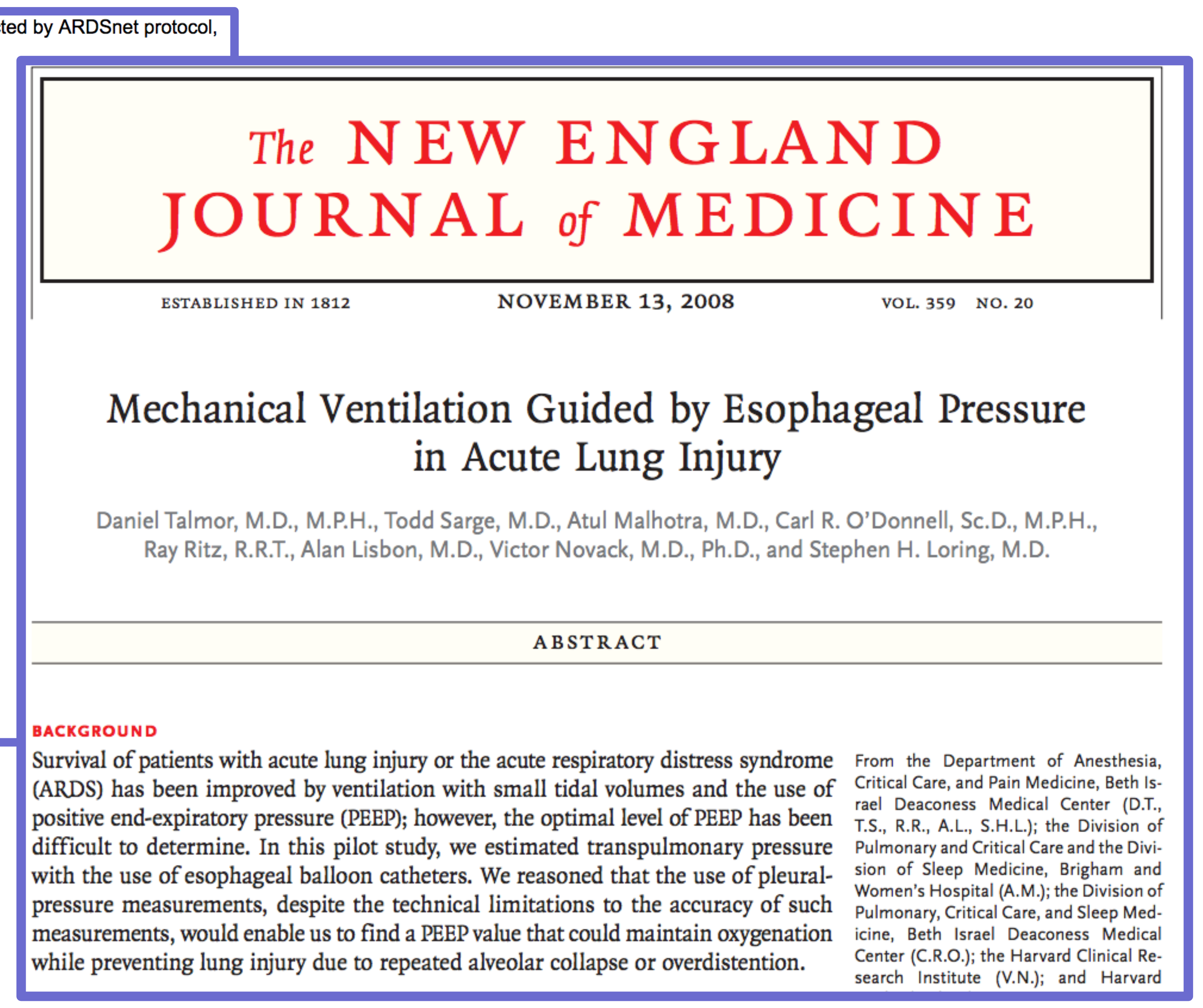
Exclusion Criteria:

- Patients with esophageal varices
- Patients with esophageal trauma
- Patients with recent esophageal surgery
- Patients with coagulopathy (platelets < 80k or International Normalized Ratio [INR]> 2)
- Post transplant patients
- Patients with significant broncho-pleural fistula

Arms

Experimental: EP
Ventilation directed by esophageal balloon measurements to maintain a positive trans-pulmonary pressure at end expiration

Active Comparator: Control
Ventilation based on the ARDSnet low tidal volume study



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 NOVEMBER 13, 2008 VOL. 359 NO. 20

Mechanical Ventilation Guided by Esophageal Pressure in Acute Lung Injury

Daniel Talmor, M.D., M.P.H., Todd Sarge, M.D., Atul Malhotra, M.D., Carl R. O'Donnell, Sc.D., M.P.H., Ray Ritz, R.R.T., Alan Lisbon, M.D., Victor Novack, M.D., Ph.D., and Stephen H. Loring, M.D.

ABSTRACT

BACKGROUND

Survival of patients with acute lung injury or the acute respiratory distress syndrome (ARDS) has been improved by ventilation with small tidal volumes and the use of positive end-expiratory pressure (PEEP); however, the optimal level of PEEP has been difficult to determine. In this pilot study, we estimated transpulmonary pressure with the use of esophageal balloon catheters. We reasoned that the use of pleural-pressure measurements, despite the technical limitations to the accuracy of such measurements, would enable us to find a PEEP value that could maintain oxygenation while preventing lung injury due to repeated alveolar collapse or overdistention.

From the Department of Anesthesia, Critical Care, and Pain Medicine, Beth Israel Deaconess Medical Center (D.T., T.S., R.R., A.L., S.H.L.); the Division of Pulmonary and Critical Care and the Division of Sleep Medicine, Brigham and Women's Hospital (A.M.); the Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center (C.R.O.); the Harvard Clinical Research Institute (V.N.); and Harvard

MIMIC-III cohort: 92 patients