CIMPOD Conference

Causal Inference Methods for PCOR using Observational Data

February 27-28, 2017 Bethesda, MD

"*Discover New Methods, Answer Patient-Centered Questions*"

Report on CIMPOD 2017 Workshop Proceedings:

Put Methods into the Practice

This Program was Funded Through a Patient-Centered Outcomes Research Institute (PCORI) Eugene Washington PCORI Engagement Award (865-MTPPI)

Held at the National Institutes of Health, February 27 - 28, 2017

CIMPOD2017.ORG

Contents

Conference Organizers

CIMPOD 2017 workshops were organized by Medical Technology & Practice Patterns Institute (MTPPI), a nonprofit organization established in 1986 to conduct research on the clinical and economic implications of health care technologies. We specialize in using electronic health records data and advanced analytical methods to conduct 'real time', 'real world' studies that are both affordable and useful for decision making. For more information email info@mtppi.org or visit our website at www.mtppi.org

MTPPI Conference Team: Yi Zhang (Project PI), Mae Thamer, Onkar Kshirsagar, and Dennis Cotter

MEDICAL TECHNOLOGY

Conference Sponsors:

IBM

MERCK

STATA®

CORECOMPETE

DELIVER VALUE FROM BIG DATA

ICSA 泛華統計協會

iClinical

Acknowledgement

CIMPOD 2017 organizers would like to thank NIDDK for its support in hosting this conference, with special thanks to Dr. Kenneth J. Wilkins, NIDDK, for his hard work on reserving NIH Natcher Conference Center rooms, coordinating with room set up, and providing support for Cisco WebEx

Context

About the workshop

The kinds of health questions you ask and the analytic methods you use to answer them have dramatically changed. Questions are now often complex and the answers require rigorous analysis using ‘causal’ modeling techniques. At CIMPOD 2016, attendees learned from internationally recognized causal methods experts and prominent stakeholder representatives how to choose wisely among various causal inference (C() techniques, focusing on principles and conceptual frameworks for selecting appropriate CI methods.

Building upon CIMPOD 2016, the theme of CIMPOD 2017 was ‘learning-by-doing' -- by attending for CIMPOD 2017’s case-study driven interactive workshops, participants were given the opportunity to learn 4 different causal inference techniques over the course of the 2-day conference. Instructors utilized real-world examples to demonstrate how these powerful methods can be implemented and utilized effectively in various research settings. CIMPOD 2017 has been a unique opportunity for researchers and patient stakeholders to learn cutting-edge and practical causal methods from a roster of internationally recognized experts. Learning Objectives included:

- Articulate your CER question effectively and select a specific causal inference method that is most appropriate for your research question
- Step-by-step implementation of causal methods with related software and program code
- Hands-on experience to help tackle the challenges related to applying the selected causal inference methods

Speakers

Greg Germino, MD

• Deputy Director of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at the National Institutes of Health (NIH)

• Prior to joining NIDDK, Dr. Germino was a full professor at the Johns Hopkins University where he directed the Johns Hopkins Polycystic Kidney Disease (PKD) Center

• CIMPOD 2017: Welcoming remark

Jason Gerson, PhD

- Senior Program Officer for the Clinical Effectiveness and Decision Science program at the Patient-Centered Outcomes Research Institute (PCORI)
- Provides intellectual and organizational leadership in designing and implementing new CER methods initiatives
- Leads some of PCORI's open science initiatives, including the development of a data sharing policy.
- CIMPOD 2017: Opening Presentation: PCORI Support of Causal inference Research: A Match(ing Methods) Made in Heaven

John Seeger, PharmD, DrPH

- Pharmacoepidemiologist and Chief Scientific Officer at Optum Epidemiology
- Conducted many studies that have addressed regulatory drug safety issues
- Worked extensively with propensity scores
- Teaches several courses on propensity scores in pharmacoepidemiology
- CIMPOD 2017 Workshops: Propensity Score

Kunjal Patel, DSc, MPH

- Senior Research Scientist, Harvard T.H. Chan School of Public Health
- Research scientist for the Pediatric HIV/AIDS Cohort Study (PHACS) and the International Maternal Pediatric Adolescent AIDS Clinical Trials group (IMPAACT)
- Extensive methodological expertise in the design and analysis of studies conducted using large observational cohorts
- CIMPOD 2017 Workshops: IPW for Static interventions

Michael Rosenblum, PhD

• Professor of Biostatistics at Johns Hopkins Bloomberg School of Public Health

• Research focuses on causal inference and development of methods for the design and analysis of randomized trials.

• Collaborates with clinical investigators in stroke, HIV prevention, cardiac resynchronization devices, and Alzheimer's disease prevention

• CIMPOD 2017 Workshops: Doubly Robust Estimators focusing on TMLE

Sonja Swanson, PhD

• Assistant Professor in the Department of Epidemiology at Erasmus MC; adjunct affiliation with the Department of Epidemiology at the Harvard T. H. Chan School of Public Health

• Methodological research focuses on improving the use and transparency of methods for estimating causal effects in epidemiology.

• CIMPOD 2017 Workshops: IV Methods

Lauren Cain, PhD

• Principal Statistician at Takeda Pharmaceuticals and a Visiting Scientist at Harvard T.H. Chan School of Public Health

• Research interests include both the development and application of methods for causal inference from complex longitudinal data

• Most recent work focuses on the optimal use of treatments through the comparison of dynamic treatment strategies.

• CIMPOD 2017 Workshops: IPW for Dynamic Interventions

Rhian Daniel, MA MSc PhD

• Assistant Professor in Biostatistics, London School of Hygiene and Tropical Medicine, Sir Henry Dale Fellow

• Interested in statistical methods for mediation analysis, time-dependent confounding and other aspects of causal inference

• Passionate about teaching and disseminating these topics to a wider audience

• CIMPOD 2017 workshops: Counterfactual-based Mediation Analysis

Jessica Young, PhD

• Assistant Professor and Biostatistician in the Department of Population Medicine, Harvard Medical School and the Harvard Pilgrim Health Care Institute

• Research focuses on the theoretical development and application of causal inference methods

• Also focuses on increasing applied researchers' access to these methods and ideas through education and software

• CIMPOD 2017 Workshops: Use of Parametric G-formula

Sherri Rose, PhD

• Associate Professor in the Department of Health Care Policy at Harvard Medical School

• Research centered around developing and integrating innovative statistical approaches to advance public health and health care research

• Methodological research focuses on nonparametric machine learning for causal inference and prediction

• Coauthored the book "Targeted Learning: Causal Inference for Observational and Experimental Data" by Springer Series in Statistics.

• CIMPOD 2017 Workshops: Machine Learning

David Drukker, PhD

• Executive Director of Econometrics at Stata

• Developed many Stata commands for estimating treatment effects

• Played a key role in the initial development of Stata MP

• Helped integrate Mata into Stata and develop some of Stata & numerical techniques

• Published on econometric methods and been principal investigator on two large research grants

• Current research interests are causal inference and spatial econometrics

• CIMPOD 2017 Workshop: Estimating Treatment Effects in Stata

Michal Rosen-Zvi, PhD

• Director for Health Informatics at IBM Research. She is also heading the Health Informatics Department at IBM Research, Haifa

• Contributed to and led a number of multidisciplinary projects where physicians, data scientists and experts from Pharmaceutical companies joined forces to analyze post launch patients data.

• Published close to forty peer-reviewed papers and co-chaired a dozen of workshops in the area of machine learning and health informatics

• CIMPOD 2017 Workshop: Observational data: Shifting the paradigm from RCTs to retrospective studies

Miguel Hernan, MD, ScM, DPH

• Professor of Department of Epidemiology and Department of Biostatistics, Harvard T.H. Chan School of Public Health

• Help organized both CIMPOD 2016 and 2017

• Conducts research on how to do research

• Develop and apply causal inference methods to guide policy and clinical interventions

• Investigates the optimal way to treat and prevent HIV infection, cancer, and cardiovascular disease

• Author of upcoming textbook "Causal Inference" (Chapman & Hall/CRC, 2013), drafts of selected chapters are available on his website

• CIMPOD 2017 Keynote Closing Address: Putting It All Together

Agenda

Day 1 Agenda

8:00-9:00	**Registration and Breakfast**
9:00-9:05	**Conference Introduction - Yi Zhang, MTPPI**
9:05-9:10	**Welcoming Remarks - Greg Germino, NIDDK Deputy Director**
9:10-9:30	**PCORI Support of Causal Inference Research: A Match(ing Methods) Made in Heaven - Jason Gerson, PCORI**
9:30-12:30	**A. Propensity Score - John Seeger, Optum** *A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction*
9:30-12:30	**1B. Constructing Inverse-Probability Weights for Static Interventions - Kunjal Patel, Harvard School of Public Health** *Long-term effectiveness of highly active antiretroviral therapy on the survival of children and adolescents with HIV infection: a 10-year follow-up study*
9:30-12:30	**1C. Doubly Robust Estimators, with Focus on Targeted Maximum Likelihood Estimation (TMLE) - Michael Rosenblum, Johns Hopkins University** *The risk of virologic failure decreases with duration of HIV suppression, at greater than 50% adherence to antiretroviral therapy*
9:30-12:30	**1D. Instrumental Variable (IV) Methods - Sonja Swanson, Erasmus Medical Center** *Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening*
12:30-1:30	**LUNCH**
1:30-4:30	**1E. Constructing Inverse-Probability Weights for Dynamic Interventions - Lauren Cain, Takeda Pharmaceuticals** *When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data*
1:30-4:30	**1F. Counterfactual-based Mediation Analysis - Rhian Daniel, London School of Hygiene and Tropical Medicine** *Causal mediation analysis with multiple mediators*
1:30-4:30	**1G. Use of the Parametric G-formula to Estimate the Effects of Time-varying Treatments - Jessica Young, Harvard Medical School** *Changes in fish consumption in midlife and the risk of coronary heart disease in men and women*
1:30-4:30	**1H. Machine Learning - Sherri Rose, Harvard School of Public Health** *Mortality risk score prediction in an elderly population using machine learning*
1:30-4:30	**1I. Estimating Treatment Effects in Stata - David Drukker, Stata**

KEEP THE CONVERSATION GOING BY USING:
TWITTER HANDLE @CIMPOD and HASHTAG #CIMPOD2017

Day 2 Agenda

8:00-9:00	**Registration and Breakfast**
9:00-12:00	**2A. Propensity Score (PS) - John Seeger, Optum** *A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute MI (further discussion)*
9:00-12:00	**2B. Constructing Inverse-Probability Weights for Static Interventions - Kunjal Patel, Harvard School of Public Health** *Atazanavir exposure in utero and neurodevelopment in infants*
9:00-12:00	**2C. Doubly Robust Estimators, with Focus on Targeted Maximum Likelihood Estimation (TMLE) - Michael Rosenblum, Johns Hopkins University** *Safety and efficacy of minimally invasive surgery plus alteplase in intracerebral haemorrhage evacuation (MISTIE): a randomised, controlled, open-label, phase 2 trial*
9:00-12:00	**2D. Instrumental Variable (IV) Methods - Sonja Swanson, Erasmus Medical Center** *Methodological considerations in assessing the effectiveness of antidepressant medication continuation during pregnancy using administrative data*
9:00-12:00	**2I. Observational Data: Shifting the Paradigm from RCTs to Retrospective Studies - Michal Rosen-Zvi, IBM Researh**
12:00-1:00	**LUNCH**
1:00-4:00	**2E. Inverse-probability Weighting (IPW) for Dynamic Interventions - Lauren Cain, Takeda Pharmaceuticals** *When to monitor CD4 cell count and HIV RNA to reduce mortality and AIDS-defining illness in virologically suppressed HIV-positive persons in high-income countries: A prospective observational study*
1:00-4:00	**2F. Counterfactual-based Mediation Analysis - Rhian Daniel, London School of Hygiene and Tropical Medicine** *How much do tumor stage and treatment explain SES inequalities in breast cancer survival? Applying causal mediation analysis to population-based data*
1:00-4:00	**2G. Use of the Parametric G-formula to Estimate the Effects of Time-varying Treatments - Jessica Young, Harvard Medical School** *Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula*
1:00-4:00	**2H. Machine Learning - Sherri Rose, Harvard School of Public Health** *A Machine Learning Framework for Plan Payment Risk Adjustment*
4:00-4:30	**Closing Keynote Address: Putting It All Together - Miguel Hernán, Harvard School of Public Health**

JOIN THE CIMPOD CONFERENCE MAILING LIST TO RECEIVE CAUSAL INFERENCE NEWS AND UPDATES: CIMPOD@MTPPI.ORG

Opening Remarks

PCORI Support of Causal inference Research: A Match(ing Methods) Made in Heaven

J A S O N G E R S O N

Following welcome and introductions, Dr. Gerson discussed why “methods matter” and how PCORI is working hard to ensure methodological rigor in the projects that PCORI support. Specifically, Dr. Gerson emphasized the following PCORI 's efforts to strengthen research methods:

- **Identify** methodological gaps relevant to the conduct of PCOR
- **Fund high impact studies** which address these gaps by developing and improving methods for research that are responsive to the needs of patients and other stakeholders
- **Disseminate** and facilitate the adoption of **new methods** to improve the conduct of PCOR

  The PCORI website contains more information on PCORI's work to promote strong methodology standards and PCORI-funded projects to accelerate PCOR and methodological research.

Dr. Gerson's presentation slides can be at CIMPOD2017.org

Workshops (A-I)

Workshop learning materials including case studies, lecture notes, recordings, data examples, and other workshop-specific items are available online at cimpod2017.org

A. Propensity Score - John Seeger, Optum

This workshop on propensity scores was built around a comparative effectiveness question involving statins. Referring to the research involved in the following reference (Am J Cardiol 2003;92:1447–1451), this workshop aimed to provide attendees with propensity score tools that can be applied to a wide range of causal questions. The research question was discussed and translated into analytic approaches, and the propensity score tool is applied to draw an answer to the research question from an observational data source. Both background and dataset with SAS programs were provided.

Case Study: A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction

Key insights

- Start with well-formed question
- Propensity score (PS) matching can achieve balance on component variables, but may not achieve balance on variables not included. Inclusion of variables to match on should be guided by a-priori, empiric approach, healthcare utilization, clinical trial results, and expert opinion
- PS not a single method. Different use of PS may produce different results. Therefore, PS analysis should conducted (and reported) in different ways including matching/stratification/weighting/modeling/restriction.
- PS method is not a panacea. Propensity matched cohorts might mimic clinical trial paradigm, but is not a randomized trial. Unmeasured covariates may not be balanced.
- We should always seek enriched data (e.g. linkages with clinical/patient data) for PS applications.
- Sensitivity analyses and supplemental data collection are often required
- PS method will NOT solve a flawed study design such as improper comparator, differential follow-up, and immortal person-time

B. Constructing Inverse-Probability Weights for Static Interventions - Kunjal Patel, Harvard School of Public Health

This workshop briefly introduced inverse probability weighting (IPW) as a method to appropriately adjust for confounding and selection bias by time-varying covariates affected by prior exposure. The primary focus of the workshop was on presenting applications of this method in clinical research. Two case studies were presented: one evaluating the effect of combination antiretroviral therapy on mortality among perinatally HIV-infected youth, and the other evaluating the effect of in-utero exposure to atazanavir on neurodevelopment among perinatally exposed, but uninfected infants. Guided exercises using SAS were conducted to help participants learn how to construct and apply IPW.

Case Study 1: Long-term effectiveness of highly active antiretroviral therapy on the survival of children and adolescents with HIV infection: a 10-year follow-up study
Case Study 2. Atazanavir exposure in utero and neurodevelopment in infants.

Key insights:

- Well-defined causal inference questions can be mapped into a target trial specified with the following components: eligibility criteria, treatment strategies, randomized treatment assignment, follow-up period, outcome, causal contrast of interest, and analysis plan.
- Conditional exchangeability assumption is required to identify a causal effect using observational data. However, this assumption is untestable. We can use expert knowledge to enhance plausibility of the assumption and measure as many relevant pre-exposure covariates as possible, but can only hope that the assumption is approximately true (i.e., there may be confounding due to unmeasured factors)
- All analytical methods assumes conditional exchangeability. *Choice of methods depend on type of treatment strategies.* All methods work for comparison of strategies involving point interventions if all baseline confounders are measured. However, advanced analytical methods (eg. IPW) are needed for comparison of sustained strategies because of time-dependent confounding and selection bias.

C. Doubly Robust Estimators, with Focus on Targeted Maximum Likelihood Estimation (TMLE) - Michael Rosenblum, Johns Hopkins University

These workshops gave two demonstrations of targeted maximum likelihood estimation. The first demonstration involved a cohort study of marginally housed HIV infected adults in San Francisco, where we estimated the causal effect of adherence to antiretroviral therapy. The second involved a randomized trial of a new surgical treatment for stroke, where we adjusted for prognostic baseline variables to improve precision and deal with informative censoring of the outcome. Dr. Rosenblum explained the double robustness property and demonstrated software implementing a targeted maximum likelihood estimator (which is double robust).

Case Study 1: The risk of virologic failure decreases with duration of HIV suppression, at greater than 50% adherence to antiretroviral therapy
Case Study 2: Safety and efficacy of minimally invasive surgery plus alteplase in intracerebral haemorrhage evacuation (MISTIE): a randomised, controlled, open-label, phase 2 trial

Key Insights:

- Definition of causal effects must be clear
- Always keep in mind the assumptions needed to identify causal effects from observed data distribution: consistency, no unmeasured confounding factors, and positivity
- Estimation methods for causal effects include: standardization, IPW, and double robust estimator. These methods are equipped to account for measured confounding
- Using bootstrap to compute Standard Errors (using bootstrap)
- Main challenges are: 1) Very small estimated values of P(Z = zjX); called “practical Experimental Treatment Assignment violation", which can lead to very large weights. May need to truncate weights; or can modify the quantity being estimated; 2) Too many variables to adjust for and not enough participants. Need to watch out for model overfit; 3) Assumption Violations (which can be hard or sometimes impossible to detect)

D. Instrumental Variable (IV) Methods - Sonja Swanson, Erasmus Medical Center

This workshop introduced instrumental variable (IV) methods, a set of methods that can potentially estimate causal effects even when confounding is unmeasured. Two case studies were presented. The first will demonstrated an IV analysis to adjust for non-compliance in a randomized trial; the second demonstrated an IV analysis to adjust for (measured and unmeasured) confounding in an observational study. In both settings, Dr. Swanson discussed the strengths and limitations of IV approaches, and applied tools for evaluating the validity and robustness of IV effect estimate

Case Study 1: Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening
Case Study 2: Methodological considerations in assessing the effectiveness of antidepressant medication continuation during pregnancy using administrative data

Key insights:

- IV methods require strong, untestable assumptions (i.e. three IV conditions for bounding; three IV conditions plus additional conditions for point estimation)
- Applying IV methods requires concerted efforts to attempt to falsify assumptions and quantify possible biases
- Under these key conditions, IV methods offer opportunities for estimating per-protocol effects in randomized trials and treatment effects in observational studies
- Transparent reporting is a key component of PCOR. Major themes in reporting guidelines apply to both IV and non-IV studies. We should always clearly state and discuss assumptions as well as the effect we are estimating.
- IV reporting also needs to address unique challenges. Seemingly minor violations of assumptions can result in large or counterintuitive biases. Identification of potential violations requires applying different subject matter expertise

E. Constructing Inverse-Probability Weights for Dynamic Interventions - Lauren Cain, Takeda Pharmaceuticals

These workshops extended the tools learned in the "Constructing Inverse Probability Weights for Static Interventions" workshop to dynamic interventions . Dynamic interventions are treatment strategies that depend on the evolution of one or more time-dependent covariates. They are regularly used in clinical practice, but rarely compared in clinical research. Two case studies comparing dynamic strategies for the care of HIV-infected individuals were presented. The first focused on the initiation of antiretroviral therapy, the other on the monitoring of biomarkers. Guided exercises using SAS provided participants hands-on experience building and implementing IPW for dynamic strategies.

Case Study 1: When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data.
Case Study 2: When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study

Key insights:

- It's an useful approach to answer a clinical question in which a target trial is described in detail and emulated. This framework can be applied to a wide variety of questions, data sources, and methods, It's advantages are 4-folds: 1) Well-defined strategies and effect estimates; 2) Avoids common biases; 3) Allows systematic evaluation; and 4) Helps explain differences between studies
- For the use of IPW to estimate effects of dynamic treatment strategies,
    - Save computing time by fitting weight model before making replicates
    - Pay careful attention to the reasons for censoring and assign contributions to the weights accordingly
    - Don't automatically stabilize weights

F. Counterfactual-based Mediation Analysis - Rhian Daniel, London School of Hygiene and Tropical Medicine

Upon finding an important socio-economic disparity in breast cancer survival, we may wish to investigate how much of this disparity is via choice of treatment. This is an example of mediation analysis. Traditionally, such a question was addressed informally and only by fitting a series of simple linear regression models. But thanks to the recent prolific contributions of VanderWeele, Vansteelandt and others, the statistical toolbox for such analyses has been hugely expanded, and the theoretical underpinning formalised. In these workshops Dr. Daniel introduced the counterfactual-based approach to mediation analysis, focusing on two case studies, in alcohol-related cardiovascular disease and breast cancer.

Case Study 1: Causal mediation analysis with multiple mediators
Case Study 2: How much do tumor stage and treatment explain socioeconomic inequalities in breast cancer survival? Applying causal mediation analysis to population-based data

Key insights:

- Questions concerning mediation are often posed and tie in with our intuition on what it means to 'understand mechanism'.
- Mediation analysis, although intuitive and with a long history, is a surprisingly subtle business as soon as there are any non-linearities in the picture.
- Advances thanks to the field of causal inference have greatly clarified these subtleties, giving rise to clear estimands that capture the notions of direct and indirect effects, clear assumptions under which these can be identified, and flexible estimation methods. However, this endeavour has been limited by *the extremely strong and untestable cross-world assumption*.
- This has effectively prohibited flexible multiple mediation analyses, even though applied problems frequently involve multiple mediators.
- Interventional effects are perhaps the way forward, since they don't require this cross-world assumption
- See Tyler VanderWeele's (2015) wonderful book for the many more topics related to mediation analysis: semiparametric estimation methods, time-to-event outcomes, three- and four-way decompositions, etc

G. Use of the Parametric G-formula to Estimate the Effects of Time-varying Treatments - Jessica Young, Harvard Medical School

In this workshop, participants learned about the parametric g-formula, an approach to estimating the effects of time-varying treatment effects using observational data with complex time-varying confounding. Dr. Young reviewed

the motivation behind the approach and the mechanics of the procedure, along with its practical advantages and disadvantages. Dr. Young then demonstrated how to use the GFORMULA SAS macro to implement this procedure in practice. Participants walked through two examples with simulated data motivated by published applications of this method. For the applied part of the workshop, attendees were required to have working knowledge of SAS and, preferably, the SAS macro language. Attendees were also asked to review the documentation for the SAS macro prior to attending the workshop. This may be accessed at https://www.hsph.harvard.edu/causal/software/.

Case Study 1: Changes in fish consumption in midlife and the risk of coronary heart disease in men and women
Case Study 2: Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula

Key insights:

- *Disadvantages of parametric g-formula*
  - Relies heavily on parametric models and subject to related bias
  - Some model misspecification can be theoretically guaranteed when null of no treatment effect is true
    - "null paradox" (Robins and Wasserman, 1997
- *Advantages of parametric g-formula*
  - More stable than other methods for continuous exposures and given "near positivity violations"
    - Occurs when an intervention level of exposure is unlikely for certain observed confounder histories
    - Parametric g-formula handles by heavier reliance on extrapolation
  - Generally, the complexity of algorithm is the same for any choice of g
    - Very little change for complex dynamic rules

H. Machine Learning - Sherri Rose, Harvard School of Public Health

Machine learning methods are most commonly used for prediction research questions, but can also be central in causal inference methods. These sessions led by Dr. Rose focused on understanding ensembled machine learning using the SuperLearner R package. This framework allows investigators to run multiple algorithms (eliminating the need to guess beforehand which single algorithm might perform best in a given data) with the opportunity to outperform any single algorithm by additionally considering all weighted averages of algorithms. The workshop also described how to incorporate the super learner within targeted maximum likelihood estimation for causal effects in the tmle R package

Case Study 1: Mortality risk score prediction in an elderly population using machine learning
Case Study 2: A Machine Learning Framework for Plan Payment Risk Adjustment

Key Insights:

- Roadmap for effect Estimation
  - Define the Research Question (I Specify Data, Specify Model, Specify the Parameter of Interest)
  - Estimate the Target Parameter
  - Inference (standard Errors / CIs; interpretation)
- Machine learning aims to "smooth" over the data and make fewer assumptions
- Key concepts for prediction include loss-based estimation, cross validation (ensembling), and flexible estimation
- Super Learner allows researchers to use multiple algorithms to outperform a single algorithm in nonparametric statistical models. It builds weighted combination of estimators where weights are optimized based on loss-function specific cross-validation to guarantee best overall fit
  - Due to its theoretical properties, super learner performs asymptotically as well as the best choice among the family of weighted combinations of estimators.

1I. Estimating Treatment Effects in Stata - David Drukker, Stata

Dr. Drukker reviewed treatment-effect estimation with observational data and discussed Stata examples that illustrated syntax and parameter interpretation. After reviewing the potential-outcome framework, the talk discussed estimators for the average treatment effect (ATE) that require exogenous treatment assignment and some estimators that allow for endogenous treatment assignment. The talk also discussed checks for balance, checks for overlap, and some estimators for the ATE from survival-time data. Finally, the talk discussed estimating and interpreting quantile treatment effects.

2I. Observational Data: shifting the paradigm from randomized clinical trials to retrospective studies - Michal Rosen-Zvi, IBM Research

Dr. Rosen-Zvi discussed the paradigm shift from RCTS to observational studies and from associational analysis to inferring causality based on observational data. Dr. Rosen-Zvi described and emphasized in detail the tools as well as the challenges when conducting causal inference using observational data.

Closing Remarks

Putting it all together - Miguel Hernan, Harvard T.H Chan School of Public Health

Dr. Hernan first outlined a framework for comparative effectiveness research using observational data that makes the target trial explicit. This framework channels counterfactual theory for comparing the effects of sustained treatment strategies, organizes analytic approaches, provides a structured process for the criticism of observational studies, and helps avoid common methodologic pitfall. Specifically, Dr. Hernan organized all case studies presented in the CIMPOD 2017 workshops based on the type of research questions and explained how the framework can be applied to these case studies.

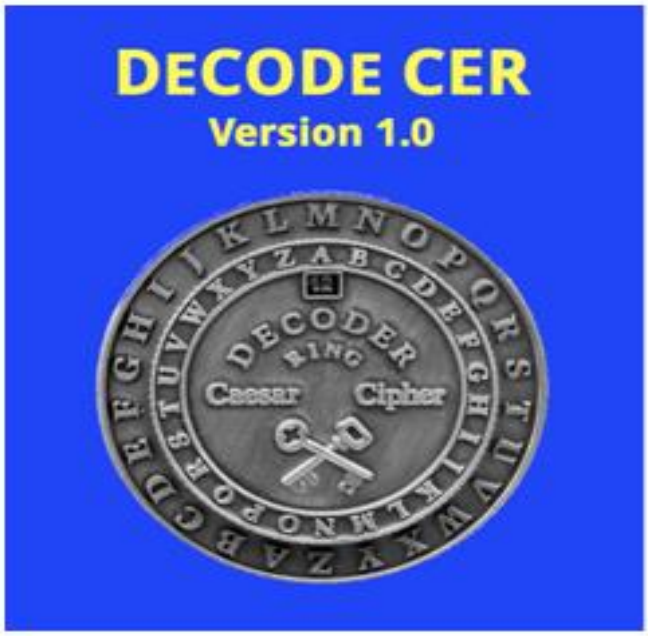
Special Presentations: Two new CER tools funded by PCORI

Announcing the Feb. 2017 Release of CERBOT.ORG— A New Web-based Tool for CER

This tool will help you design your CER study using causal inference (CI) statistical methods. By navigating the five modules found in CERBOT, you will 'emulate' a randomized clinical trial using your observational data source. CERBOT walks you through – step by analytic step – the entire process of designing your CER study and selecting the suitable causal inference methods. *Features include*:

- Personalized Account
- Articulation of Research Questions
- Research Team Collaboration through Selection of Stakeholders
- Downloadable Reports
- Detailed Case Study References
- CI Methods Recommendations Based on Your Data and Question

CERBOT was Funded Through a Patient-Centered Outcomes Research Institute (PCORI) Improving Methods for Conducting CER Award (ME-1303-6031) Principal Investigators: Yi Zhang, MTPPI and Miguel Hernan, Harvard T.H. Chan School of Public Health



The 'Decision Tool for Causal Inference and Observational Data Analysis Methods in Comparative Effectiveness Research' (DECODE CER)

CER seeks to compare interventions in real-world settings. Causal inference methods have become increasingly popular for CER, but debate exists on the utility of different approaches. We developed a decision tool to aid researchers in formulating questions, recognizing assumptions, conducting analyses, and interpreting results.

Please contact Douglas Landsittel, PhD, at douglandsittel@pitt.edu for more information.