INTRODUCTION TO STATISTICAL APPROACHES TO COMPARATIVE EFFECTIVENESS RESEARCH

Sharon-Lise Normand

Harvard Medical School & School of Public Health

September 14, 2017

WEBINAR DESCRIPTION

Comparative Effectiveness Research (CER) refers to a body of research that generates and synthesizes evidence on the comparative benefits and harms of alternative interventions to prevent, diagnose, treat, and monitor clinical conditions, or to improve the delivery of health care. The evidence from CER is intended to support clinical and policy decision making at both the individual and the population level. While the growth of massive health care data sources has given rise to new opportunities for CER, several statistical challenges have also emerged. This webinar provides an overview of the types of research questions addressed by CER, reviews the main statistical methodology currently utilized, and highlights areas where new methodology is required. Inferential issues in the big data context are identified. Examples from cardiology will illustrate methodological issues.

GOALS

- Understand what comparative effectiveness research (CER) constitutes
  - Know key features
- Identify statistical methodology utilized in contemporary examples of CER
- Highlight methodological challenges

OUTLINE

- **12:00 - 12:30: Introduction**
  - Three examples
  - What is CER and why important?
  - Main characteristics of CER
- **12:30 - 1:45: Methodology**
  - Assumptions and causal parameters
  - Approaches
    - General settings
    - High dimensional settings
  - Example
- **1:45- 2:00: Closing Remarks**
  - Research needs

THANKS

- Lauren Kunz (formerly PhD student): NHLBI
- Francesca Dominici & Cory Zigler: Harvard T. H. Chan School of Public Health
- I gratefully acknowledge support from:
  - **U01-FD004493** (U.S. Food and Drug Administration)
  - **R01-GM111339** (U.S. National Institute of General Medical Sciences)

COLLABORATORS

- Sherri Rose, Harvard Medical School, Associate Professor of Health Care Policy (Biostatistics)
- Jacob Spertus, Harvard Medical School (Research Assistant)

RADIAL VS FEMORAL PCI

- Radial artery access permits easier access and easier closure
- Large number of patients undergoing both procedures
- Not particularly well studied and of growing importance in the U.S.
- Marked heterogeneity in predisposition to bleeding
- Significant treatment selection (healthier patients undergo transradial procedures)

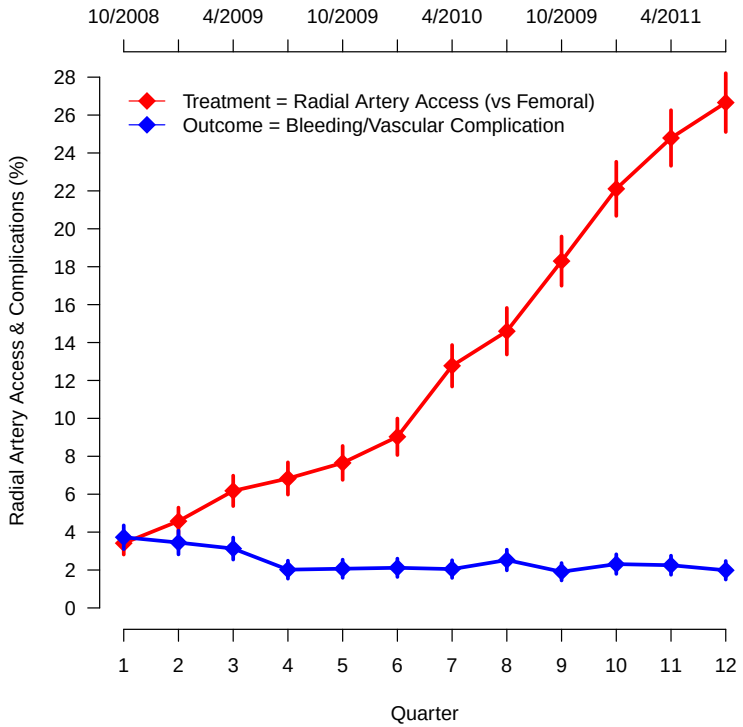
Does **radial artery access** cause fewer complications compared to **femoral artery access**?

RADIAL VS FEMORAL PCI

- Radial artery access permits easier access and easier closure
- Large number of patients undergoing both procedures
- Not particularly well studied and of growing importance in the U.S.
- Marked heterogeneity in predisposition to bleeding
- Significant treatment selection (healthier patients undergo transradial procedures)

Does **radial artery access** cause fewer complications compared to **femoral artery access**?

MASSACHUSETTS



40,000 PCIs in MA adults

Source = Mass-DAC
Kunz, Rose, et al., 2017

DRUG ELUTING (DES)VERSUS BARE METAL (BMS) CORONARY STENTS

- DES (approved 2003) and BMS (approved in 1990s) frequently implanted to keep treated arteries clear & supported
- DES improves target-vessel revascularization (TVR) more than BMS
- DES associated with late stent thrombosis (death)
- Have **9000** patients and **500** confounders

Do **DES** cause fewer revascularizations compared to **BMS**?

MASSACHUSETTS, 2011

	Stent Type	
Characteristic	BMS	DES
Outcomes, %		
1 Year Mortality	**10.2**	**3.3**
1 Year TVR	9.0	6.5
Confounders		
Age, yrs	66.4	63.7
STEMI, %	35.7	18.2
Cardiomyopathy or LVSD, %	11.1	8.4
Emergent, %	38.3	20.3
Shock, %	3.8	0.8

Source = Mass-DAC

Spertus and Normand, 2017 (under review)

SPECIFIC DRUG ELUTING CORONARY STENTS

- Rapid proliferation of drug eluting stents (DES)
- U.S. has 2nd highest number of overall stent insertions per capita
- Multiple competing versions supported by a few manufacturers
- Differences include polymer coating, specific drug, platform type, and delivery system
- Study **21,000+** adults, **10** model-specific DES, **3** manufacturers

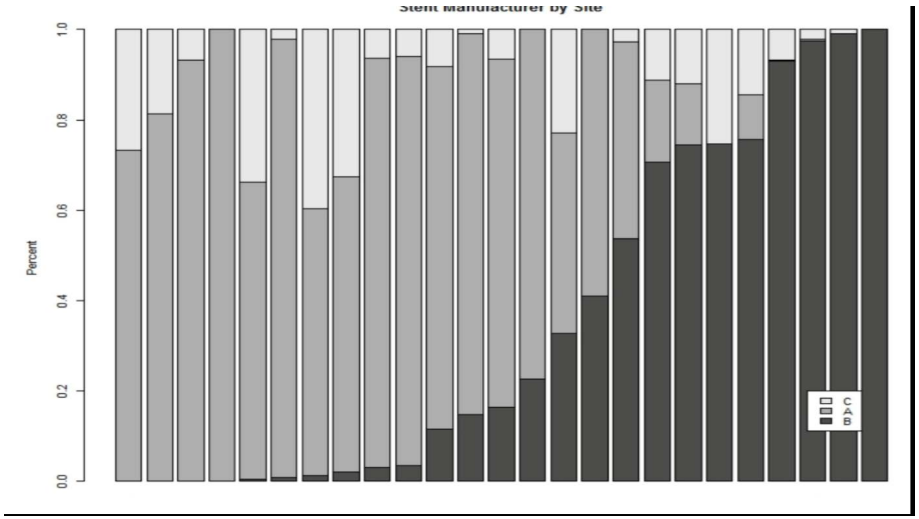
Do particular **model-specific DES** cause fewer revascularizations compared to other **model-specific DES**?

SPECIFIC DRUG ELUTING CORONARY STENTS

- Rapid proliferation of drug eluting stents (DES)
- U.S. has 2nd highest number of overall stent insertions per capita
- Multiple competing versions supported by a few manufacturers
- Differences include polymer coating, specific drug, platform type, and delivery system
- Study **21,000+** adults, **10** model-specific DES, **3** manufacturers

MASSACHUSETTS

% of DES manufacturer of all DES implanted, by hospital, 2008-2012.



Rose and Normand, 2017 (under review)
Source: Mass-DAC

Do particular **model-specific DES** cause fewer revascularizations compared to other **model-specific DES**?

WHAT IS CER?

PATIENT CENTERED OUTCOMES RESEARCH INSTITUTE DEFINITION

- Assesses benefits/harms of **preventive, diagnostic, therapeutic, palliative, or health delivery system interventions** to inform decision making using comparisons & outcomes that matter to people;
- Includes an individualś preferences, autonomy/needs, focusing on outcomes that people notice & care about (**survival, function, symptoms, and health related quality of life**);
- Incorporates a **variety of settings/diversity of participants** to address individual differences & barriers to implementation or dissemination; and
- Investigates **optimizing outcomes** while addressing **burden** to individuals, resource availability, & other stakeholder perspectives.

TENETS OF CER

1. Directly informs a specific clinical decision from **patient perspective** OR a health policy decision from **population perspective**.
2. Compares at least **two** alternatives (one could be usual care or best practice).
3. Describes results at population and **subgroup** levels.
4. Measures **benefits and harms** important to patients.
5. Employs **methods and data sources** appropriate for the decision of interest.
6. Conducted in **settings** that are similar to those in which intervention will be used in practice.

TYPOLOGY OF CER STUDIES

- Experimental
  - Randomized
    - Cluster
    - Adaptive
    - **Pragmatic:** some discussion in this webinar
  - Controlled
- Observational
  - **Prospective/retrospective:** focus of this webinar
  - Microsimulation modeling
- Research Synthesis
  - Cross-design, network, and meta-analysis
  - Decision analysis

OUTLINE

- Introduction
  - Three examples
  - What is CER and why important?
  - Main characteristics of CER
- **Methodology**
  - Assumptions and causal parameters
  - Approaches
    - General setting
    - High-dimensional setting
  - Examples
- Closing Remarks
  - Research needs

DATA, PARAMETERS, NOTATION

- $T = 1$ new and $T = 0$ standard treatment(binary)
- Y observed outcome
- Y_1, Y_0 **potential outcomes** under $T = 1$ and $T = 0$
- $\mathbf{X}$ is a set of (baseline) covariates
- Data: $(T_i, Y_i, \mathbf{X}_i), i = 1, \cdots, N$
- Mean marginal outcome under treatment:

$$\mu_1 = E_X\left(E(Y \mid T = 1, \mathbf{X})\right)$$

- Mean marginal outcome under standard treatment:

$$\mu_0 = E_X\left(E(Y \mid T = 0, \mathbf{X})\right)$$

- Interested in the marginal **effect** of T on Y

$$\Delta = \mu_1 - \mu_0 \quad \text{(Difference)} \quad \text{or} \quad \Delta = \frac{\mu_1}{\mu_0} \quad \text{(Ratio)}$$

DATA, PARAMETERS, NOTATION

(T Treatment, Y Outcomes, $\mathbf{X}$ Covariates)

- Average Treatment Effect **(ATE)**

$$\mathrm{E}(\mathrm{Y}_1 - \mathrm{Y}_0) = \mathrm{E}_{\mathrm{X}}\left(\mathrm{E}(\mathrm{Y} \mid \mathrm{T} = 1, \mathbf{X}) - \mathrm{E}(\mathrm{Y} \mid \mathrm{T} = 0, \mathbf{X})\right) \tag{1}$$

  Expected outcome change if units randomly assigned to treatment of comparison group

ATE may contain effect on subjects for whom the treatment was not intended (food voucher programs)

- Average effect of the Treatment on the Treated **(ATT)**

$$\begin{aligned}\mathrm{E}(\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{T} = 1) = \mathrm{E}_{\mathrm{X}}\{&\mathrm{E}(\mathrm{Y} \mid \mathrm{T} = 1, \mathbf{X}) \\ &- \mathrm{E}(\mathrm{Y} \mid \mathrm{T} = 0, \mathbf{X}) \mid \mathrm{T} = 1)\}\end{aligned} \tag{2}$$

  Expected outcome for a randomly selected unit from the treatment group

ASSUMPTIONS

($\mathbb{T}$ Treatment, $\mathbb{Y}$ Outcomes, $\mathbb{X}$ Covariates)

Potential Outcomes assumed to exist

The potential outcomes assumption is fundamental

1. Stable Unit Treatment Value Assumption
2. Ignorability of Treatment Assignment
3. Positivity
4. Constant Treatment Effect

If (1) & (2) are violated, then causal parameters can be estimated **statistically** but **cannot** be interpreted **causally**

ASSUMPTIONS

($\mathbf{T}$ Treatment, $\mathbf{Y}$ Outcomes, $\mathbf{X}$ Covariates)

Stable Unit Treatment Value Assumption no interference and no variation in treatment.

1. Potential outcomes for a unit do not depend on the treatment assignment of other units (no spillover effects)

$$Y_i(T_1, T_2, \cdots, T_N) = Y_i(T_i) = Y_{it} \tag{3}$$

   - **Radial artery access:** violated if as physician increases skill in radial artery access, the less likely complications arise, and the more likely the physician is to use radial access on subsequent subject.

2. Treatments are well-defined and the **same** for all units
   - **Radial artery access:** violated if physicians accessing radial artery use different methods of applying pressure after removing catheter.

ASSUMPTIONS

(T Treatment, Y Outcomes, $\mathbf{X}$ Covariates)

Ignorability of Treatment Assignment unconfoundedness of treatment assignment

1. Within subpopulations defined by $\mathbf{X}$, random treatment assignment

$$(Y_1, Y_0) \perp \mathrm{T} \mid \mathbf{X} \tag{4}$$

$$\mathrm{P}(\mathrm{T} = 1 \mid Y_1, Y_0, \mathbf{X}) = \mathrm{P}(\mathrm{T} = 1 \mid \mathbf{X}) \tag{5}$$

- Untestable assumption (sensitivity analysis, multiple comparison groups, control outcomes)
- **Radial artery access:** violated if a covariate associated with probability of undergoing radial artery access as well associated with a complication is omitted

SELECTION BIAS

($\mathbf{T}$ Treatment, $\mathbf{Y}$ Outcomes, $\mathbf{X}$ Covariates)

Suppose $\mathbf{U}$ is a vector of **unmeasured** confounders and

$$Y_{1i} = \mathbf{X}'\boldsymbol{\beta}_1 + \mathbf{U}_i'\boldsymbol{\alpha}_1 \quad (6)$$

$$Y_{0i} = \mathbf{X}'\boldsymbol{\beta}_0 + \mathbf{U}_i'\boldsymbol{\alpha}_0 \quad (7)$$

Observed outcome is

$$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i} \quad (8)$$

Substituting (6) and (7) into (8) yields

$$Y_i = \mathbf{X}'\boldsymbol{\beta}_0 + T_i \left(\underbrace{(\mathbf{X}'\boldsymbol{\beta}_1 - \mathbf{X}'\boldsymbol{\beta}_0}_{\text{X Observed}} + \underbrace{(\mathbf{U}_i'\boldsymbol{\alpha}_1 - \mathbf{U}_i'\boldsymbol{\alpha}_0)}_{\text{U Unobserved}} \right) + \mathbf{U}_i'\boldsymbol{\alpha}_0$$

SELECTION BIAS (T Treatment, Y Outcomes, $\mathbf{X}$ Observed Covariates, $\mathbf{U}$ Unmeasured Covariates)

$$\mathsf{Y}_i = \mathbf{X}'\boldsymbol{\beta}_0 + \mathsf{T}_i \left(\underbrace{(\mathbf{X}'\boldsymbol{\beta}_1 - \mathbf{X}'\boldsymbol{\beta}_0}_{\text{X Observed}} + \underbrace{(\mathbf{U}_i'\boldsymbol{\alpha}_1 - \mathbf{U}_i'\boldsymbol{\alpha}_0)}_{\text{U Unobserved}} \right) + \mathbf{U}_i'\boldsymbol{\alpha}_0$$

$$\begin{aligned} (\mathbf{U}, \mathbf{X}) &\perp \mathsf{T} && \text{randomization} \\ \mathbf{U} &\perp \mathsf{T} \mid \mathbf{X} && \text{ignorable treatment assignment} \\ \mathbf{U} &\not\perp \mathsf{T} \mid \mathbf{X} && \text{hidden bias} \end{aligned}$$

Different terminology:

- No unmeasured confounders = no omitted variables = overt bias = ignorable treatment assignment
- hidden bias = residual confounding = omitted variables = non-ignorable treatment assignment

ASSUMPTIONS

(T Treatment, Y Outcomes, $\mathbf{X}$ Covariates)

Positivity

1. Requires units at every combination of observed covariates so that probability bounded away from zero

$$1 > \mathrm{P}(\mathrm{T} = 1 \mid \mathbf{X}) > 0 \tag{9}$$

- Structural violations when units associated with specific covariate values cannot possibly get the treatment
- Practical violations due to **finite** sample size**
- Statistically testable
- **Radial artery access:** examine covariate **balance** and covariate **overlap**

CAUSAL ASSUMPTIONS (T Treatment, Y Outcomes, X Covariates)

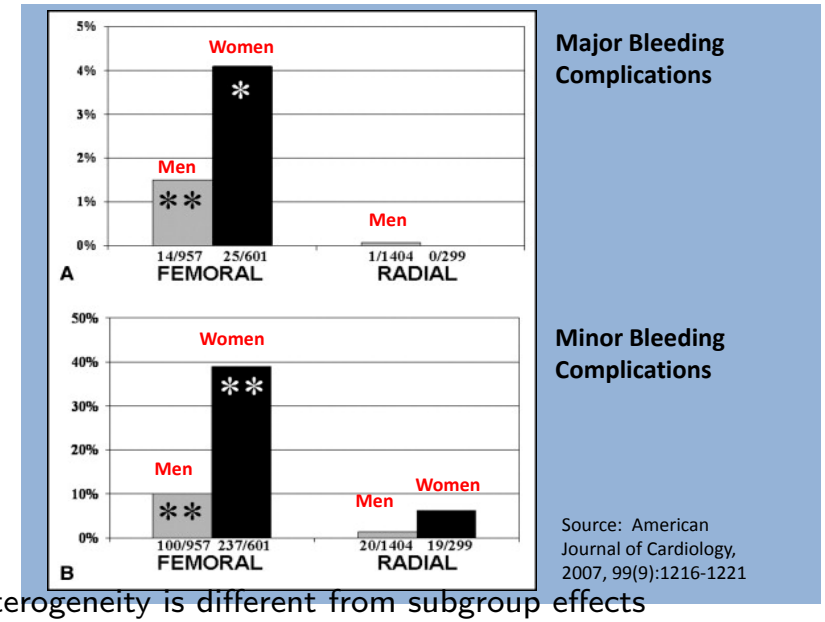
Constant Treatment Effect

1. Observable treatment effect for any two units having the same values of $\mathbf{X}$ should be similar

$$E(Y_1 - Y_0 \mid T = 1, \mathbf{X}) = E(Y_1 - Y_0 \mid T = 0, \mathbf{X})$$

- If not violated, the ATE may be interpreted both marginally and conditionally
- **Radial vs Femoral access:** between-subject differences in bleeding complications vary due to subject-specific idiosyncrasies

Radial Artery Access: women may have a bigger benefit than men



Heterogeneity is different from subgroup effects

APPROACHES

(T Treatment, Y Outcomes, $\mathbf{X}$ Covariates)

Joint Distribution

$$\begin{aligned} \mathrm{P}(Y, \mathsf{T}, \mathbf{X}) &= \mathrm{P}(Y \mid \mathsf{T}, \mathbf{X}) \times \mathrm{P}(\mathsf{T} \mid \mathbf{X}) \times \mathrm{P}(\mathbf{X}) \\ &= \mathrm{Q}_{\mathrm{Y}} \times \Pi_{\mathsf{T}} \times \mathrm{Q}_{\mathrm{X}} \end{aligned} \tag{10}$$

1. Treatment effect depends **only** on Q_{Y} **and** Q_{X}

$$\mathrm{E}_{\mathrm{X}}\left(\mathrm{E}(Y \mid \mathsf{T}=1, \mathbf{X})\right) - \mathrm{E}_{\mathrm{X}}\left(\mathrm{E}(Y \mid \mathsf{T}=0, \mathbf{X})\right) \tag{11}$$

2. Π_{T} is the **propensity score** (nuisance)

$$\Pi_{\mathsf{T}} = \mathrm{P}(\mathsf{T}=1 \mid \mathbf{X})$$

$\mathbf{X}$ could be very high-dimensional

APPROACHES: 3 TYPES (considered today) (T Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

1. Model only the treatment assignment mechanism via regression
2. Model only the outcome via regression
3. Model both the treatment assignment mechanism and outcome

WLOG: **Causal Parameter:** $\Delta = E(Y_1 - Y_0)$ (ATE)

TREATMENT ASSIGNMENT ONLY (T Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

Weight by the Inverse Probability of Treatment (IPTW)

- Horvitz-Thompson (1952):

$$\hat{\Delta}_{\text{IPTW-HT}} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i Y_i}{\Pi_{T_i}} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-T_i)Y_i}{1-\Pi_{T_i}}$$

- Rosenbaum and others (Rosenbaum, Rubin; 1984):

$$\hat{\Delta}_{\text{IPTW-R}} = \left(\sum_{i=1}^{N}\frac{T_i}{\Pi_{T_i}}\right)^{-1}\sum_{i=1}^{N}\frac{T_i Y_i}{\Pi_{T_i}} - \left(\sum_{i=1}^{N}\frac{1-T_i}{1-\Pi_{T_i}}\right)^{-1}\sum_{i=1}^{N}\frac{(1-T_i)Y_i}{1-\Pi_{T_i}}$$

OUTCOME MODEL ONLY (T Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

Regression modeling

- Multiple regression modeling (standard statistical text books)
- Parametric g-computation (Snowden, Rose, Mortimer; 2011)

$$\hat{\Delta}_{\text{G-comp}} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \hat{E}(Y \mid T_i = 1, \mathbf{X}_i) - \hat{E}(Y \mid T_i = 0, \mathbf{X}_i) \right\}$$

$\hat{E}(Y \mid T_i = t, \mathbf{X}_i)$ is the regression of Y on $\mathbf{X}$ in treatment group t

TREATMENT AND OUTCOME MODELS (T

Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

Weight by the IPTW and estimate a regression

- Augmented IPTW (Robins, Rotnitzky, Zhao; 1994): $\hat{\Delta}_{\text{A-IPTW}}$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\{I(T_i = 1) - I(T_i = 0)\}}{\Pi_{T_i}}(Y_i - \hat{E}(Y \mid T_i, \mathbf{X}_i))$$

$$+\frac{1}{N}\sum_{i=1}^{N}\left(\hat{E}(Y \mid T_i = 1, \mathbf{X}_i) - \hat{E}(Y \mid T_i = 0, \mathbf{X}_i)\right)$$

TREATMENT AND OUTCOME MODELS

(T Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

Targeted Maximum Likelihood (van der Laan, Rose; 2011), $\hat{\Delta}_{\text{TMLE}}$:

$$\frac{1}{N}\sum_{i=1}^{N}\left(E^*(Y \mid T=1, \mathbf{X_i}) - E^*(Y \mid T=0, \mathbf{X_i})\right)$$

$$E^*(Y \mid T=t, \mathbf{X}) = E^0(Y \mid T=t, \mathbf{X}) + \epsilon_t H^*(T, \mathbf{X})$$

$$H^*(T, \mathbf{X}) = \frac{T}{\Pi_T} - \frac{1-T}{1-\Pi_T}$$

- E^*: **targeted** estimate of regression of Y on $(T, \mathbf{X})$ obtained by moving the initial estimate E^0 by fluctuations defined by $\epsilon_t H^*(T, \mathbf{X})$
- **Key idea**: no need to maximize entire likelihood $(Y, T, \mathbf{X})$ because causal parameter only depends on Q_Y and Q_X

SUMMARY

Approach (R function)	Strengths	Weaknesses
IPTW (matching)	Simple	Large variance estimates Weight trimming bias
Regression (glm)	Parametric Simple	Extrapolation if violate positivity Functional form
G-Comp	Parametric Simple	Extrapolation if violate positivity Functional form
A-IPTW (twang)	Double robust Asymptotic efficiency	Finite sample inefficient
TMLE (tmle)	Double robust Asymptotic efficiency Finite sample efficiency	

WHAT ARE BIG DATA?

- Experimental units (units of primary interest): patients, physicians, hospitals, mental health centers, health plans
- (Number of experimental units)/(number of unknown parameters) $= n/k$
  - $\approx$measure of how much information is available to estimate each unknown parameter
  - Estimating the association of age on patient satisfaction (y) following surgery for a cohort of 100 surgical patients

$$y = \beta_0 + \beta_1 \times \text{age} + \text{error} \tag{12}$$

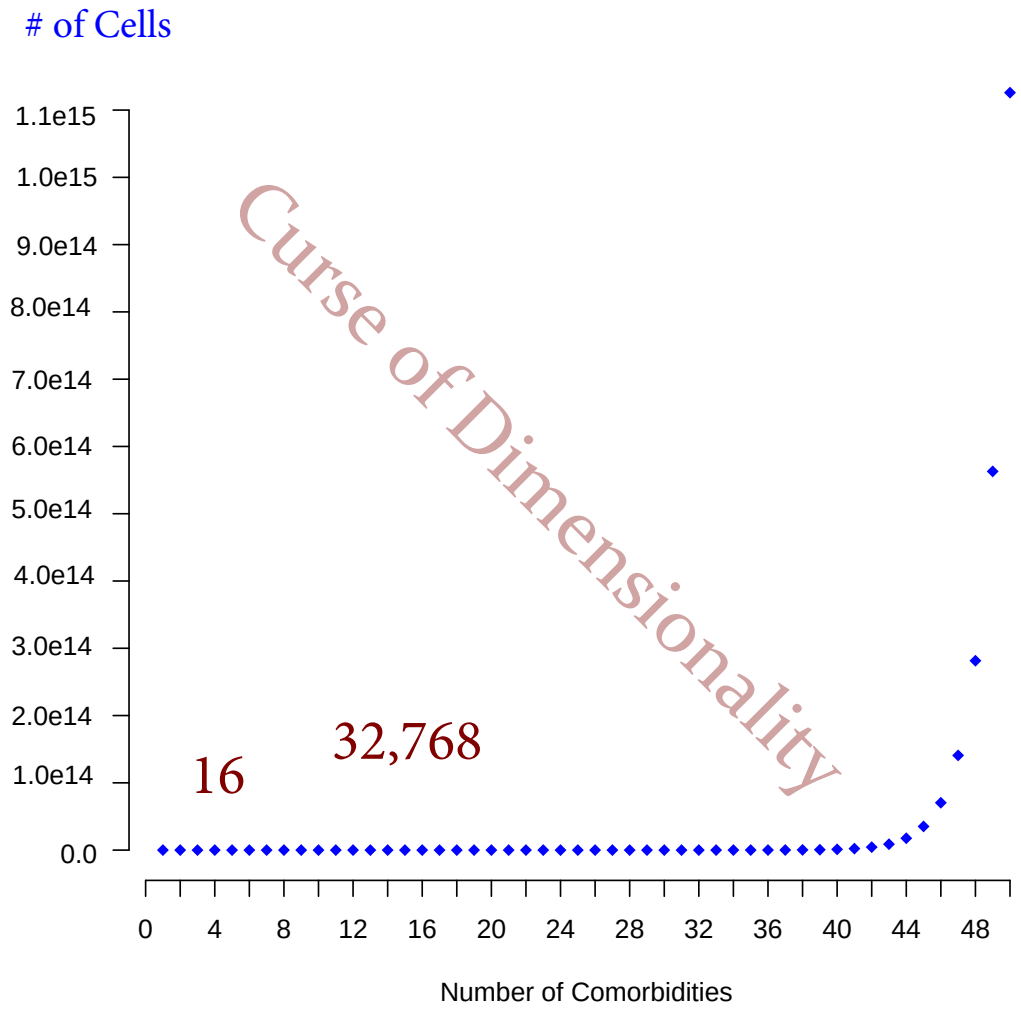
  - $100/2 = 50 \rightarrow \approx$50 independent pieces of of information to estimate each unknown parameter
- **Big data**: $k > n$ or **sparsity**

MAIN PROBLEM

- Theory based on what happens when n gets large while k is **fixed**
- **Sparsity** - not enough data to estimate all parameters of interest
- Suppose $k = 15$ binary-valued comorbidities (e.g., schizophrenia, substance use disorder, etc.)
  - Number of unique comorbidity patterns $> 32{,}000$

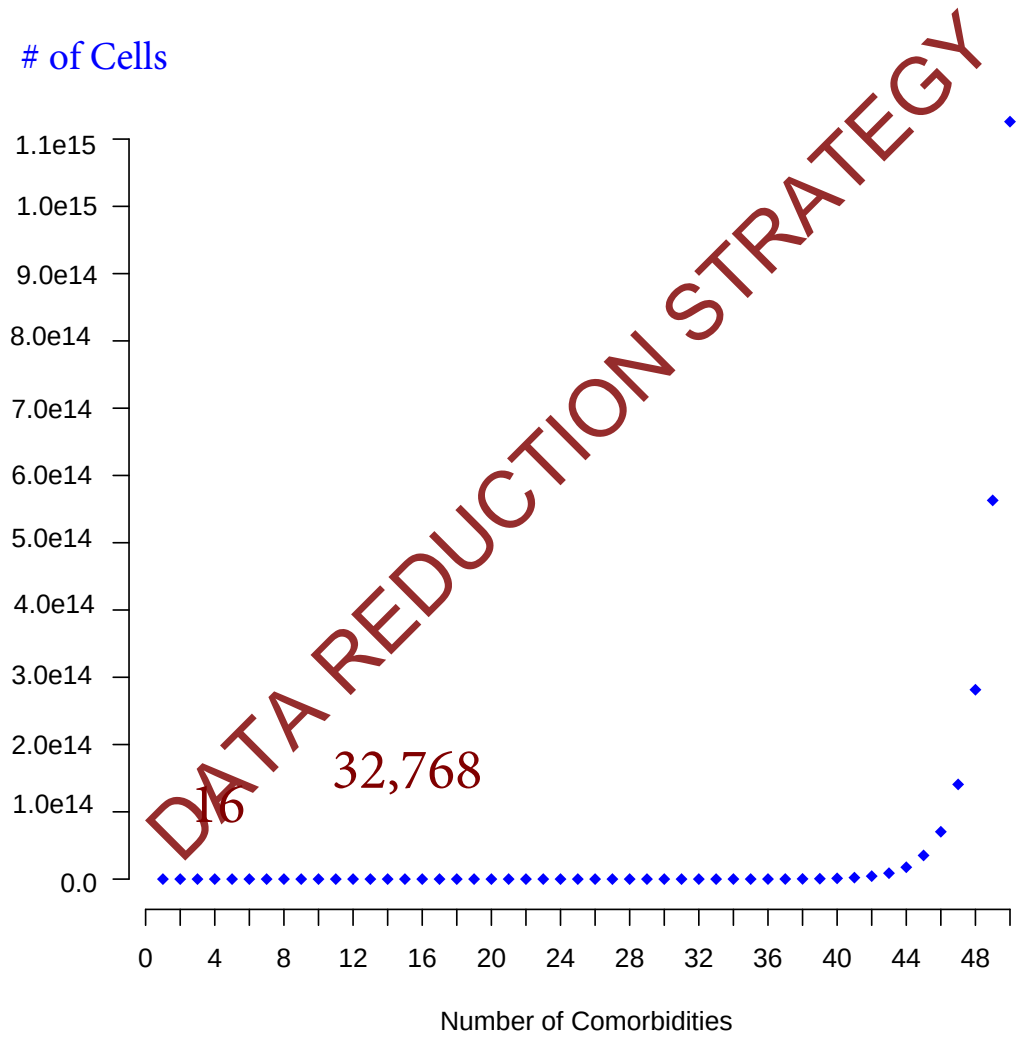
MAIN PROBLEM

- Theory based on what happens when n gets large while k is **fixed**
- **Sparsity** - not enough data to estimate all parameters of interest
- Suppose $k = 15$ binary-valued comorbidities (e.g., schizophrenia, substance use disorder, etc.)
  - Number of unique comorbidity patterns $> 32{,}000$



MAIN PROBLEM

- Theory based on what happens when n gets large while k is **fixed**
- **Sparsity** - not enough data to estimate all parameters of interest
- Suppose $k = 15$ binary-valued comorbidities (e.g., schizophrenia, substance use disorder, etc.)
  - Number of unique comorbidity patterns $> 32{,}000$



OTHER CHALLENGES

- **Overfitting** - a model is so **specific**, it reproduces idiosyncrasies of the particular data from which the model parameters were estimated
  - Will perform poorly in new data sets
- **Parametric assumptions** - with many covariates, unlikely to get the functional form correct
  - Only linear terms, splines, interactions, etc.?
- Uncertainty about subpopulations experiencing heterogeneous effects (many potential subpopulations)
- **Selective** inference - we only see results based on some selected feature (e.g., statistically significant) of the data

What inferential approaches are available?

DIMENSION REDUCTION TECHNIQUES

1. Limit number of confounders based on **perceived** clinical relevance & estimate a single model
   - Most frequently utilized
   - Implicitly a **causal inference** tool
2. **Sparseness**: assume a small number of variables represent the underlying data structure
   - LASSO (least absolute shrinkage and selection operator); find $\hat{\boldsymbol{\beta}}$ minimizes $\sum_i (Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2 + \lambda \boxed{\sum_{j=1}^{k} |\beta_j|}$
   - $\lambda =$ penalty for too many variables (large λ penalizes more complex models) selected by minimizing squared leave-one-out errors
   - Sparse additive models (SpAM): minimize

$$\sum_i (Y_i - \sum_{j=1}^{k} f_j(X_{ij}))^2 + \lambda \boxed{\sum_{j=1}^{k} \sqrt{f_j^2(X_{ij})}} + \boxed{\sum_{j=1}^{k} \mu_j f_j}$$

DIMENSION REDUCTION TECHNIQUES

3. **Denseness**: shrink estimates to a common mean & permit a small number of variables to have distinct coefficients
   - Kernel Regularized least squares (Gelman et al, 2008)
   - Key idea: information is encoded in the **similarity** between observations (small distances) such that more similar observations should have more similar outcomes

$$\textbf{Gaussian Kernel} \quad k(\mathbf{x_i}, \mathbf{x_j}) = \exp -\frac{||\mathbf{x_i} - \mathbf{x_j}||^2}{\sigma^2}$$

4. **Denseness and sparseness**: shrink estimates to a common mean and to zero so two penalty terms; Elastic Net (Zou and Hastie, JRSSB 2005)
5. **Ensemble Techniques**: combine a large number of models to obtain a stronger model (usually for **prediction**)
   - Bagging: averages the models in the ensemble (random forests)
   - Boosting: iteratively learns from weaker models (generalized boosting)
   - Stacking/SuperLearner: ensemble strong, diverse models

DIMENSION REDUCTION FOR COMPARATIVE EFFECTIVENESS

- High-dimensional **propensity score** (Schneeweiss et al., 2009)
  - Binary treatment, binary outcome, binary confounders
  - Rank confounders for inclusion
  - No accounting for **uncertainty** in variable selection
- **Target maximum likelihood** (TMLE package in R)
  - No need to maximize entire likelihood (van der Laan, Rose, 2011)
  - Semi-parametric
- **2-Step Approach**:
  - Estimate propensity score model then a **parametric** outcome model (Kaplan and Chen, 2012))
  - Saarela et al., 2015: marginal model specification coupled with inverse probability of treatment weighting 2-step procedure

DIMENSION REDUCTION FOR COMPARATIVE EFFECTIVENESS

- Bayesian **model average** + adjustment for confounding
  - Binary treatment & outcomes (Zigler and Dominici, 2014)
  - Parametric assumptions for outcome equation
  - Could use Bayesian **regularization** instead: Discrete mixtures or shrinkage priors
- Model propensity score as a latent variable jointly with outcome model (McCandless et al. 2009)

MODEL AVERAGING & ADJUSTMENT

(T Treatment, Y Outcomes, X Covariates)

Joint estimation of propensity score models and outcome models

- Assumed strongly ignorable treatment assignment

$$g(T_i) = \gamma_0 + \sum_{j=1}^{p} \left(\alpha_j^X\right) \gamma_j X_{ij}$$

$$Y_i = \beta_0^{\alpha^Y} + \beta_T^{\alpha^Y} T_i + \sum_{j=1}^{p} \left(\alpha_j^Y\right) \beta_j^{\alpha^Y} X_{ij} + \epsilon_i^Y$$

α_j^Y and α_j^X = "inclusion" probabilities

Confounders: **large** values of **both** α_j^Y & α_j^X

BAYESIAN COMPUTATIONAL APPROACH

WANT: $\Delta =$ **E(**Y_1**) - E(**Y_0**)**

Common two-step approach

1. Step 1: estimate propensity score model via **regularization**; **single** estimated score for each individual, $\hat{\pi}(\text{DES} = 1 \mid \text{confounders})$
2. Step 2: use estimated score to compute $\hat{\Delta}$ via matching, weighting, or stratification

New two-step approach

1. Step 1: same as above **except** get **distribution** of score for each individual
2. Step 2: for each draw for each individual of the propensity score, calculate the weighted risk difference $\hat{\Delta}$

REGULARIZATION VIA SHRINKAGE

(T Binary Treatment, Y Binary Outcome, X Confounders, k Large)

$$
\begin{aligned}
T_i &\sim \text{Bern}(\pi(\mathbf{X}_i)) \\
\pi(\mathbf{X}_i) &= \text{logit}^{-1}\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}\right)
\end{aligned}
$$

τ = global scale parameter

λ_j local scale parameter

Priors for β_j:

- Typically centered at 0
- Normal, double-exponential, Student-t
- Horseshoe prior (Carvahlo et al., 2010)

$$
\begin{aligned}
\beta_j &\sim \text{N}(0, \lambda_j^2 \tau^2) \\
\lambda_j, \tau &\sim \text{Cauchy}^{+}(0, 1)
\end{aligned}
$$

- Mimics Bayesian Model Averaging (with heavy-tailed discrete mixtures)

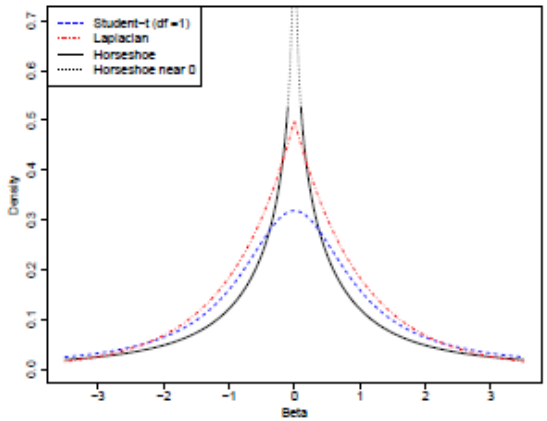


Figure 1: The horseshoe prior and two close cousins: Laplacian and Student-t.

Source: Carvalho CM, Polson NG, Scott JG. Handling Sparsity via the Horseshoe. Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, 2009.

MODELS (T Binary Treatment, Y Binary Outcome, X Confounders)

Step 1: Treatment Model

$$\begin{aligned} T_i &\sim \text{Bern}(\pi(\mathbf{X}_i)) \\ \pi(\mathbf{X}_i) &= \text{logit}^{-1}\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}\right) \end{aligned}$$

$\pi(\mathbf{X}_i)$ = propensity score

- Priors required for β_j
- Regularize via distribution: shrinkage priors (continuous priors centered on 0)

Step 2: Outcome Model

$$\begin{aligned} Y_T \mid n_T, p_T, \pi(\mathbf{X}) &\sim \text{Binomial}(n_T, p_T) \\ p_T \mid \alpha_{T0}, \alpha_{T1} &\sim \text{Beta}(\alpha_{T0}, \alpha_{T1}) \end{aligned}$$

n_T = number of subjects receiving treatment T
p_T = probability of outcome under treatment T
Independent samples in treatment groups

- *A-posteriori*

$$p_T \mid \mathbf{Y}, \pi(\mathbf{X}) \sim \text{Beta}(a_T, b_T)$$

MODEL (T Binary Treatment, Y Binary Outcome, X Confounders, k Large)

Posterior Distribution: $p_T \mid \mathbf{Y}, \boldsymbol{\pi}(\mathbf{X}) \sim \text{Beta}\,(a_T, b_T)$
$T = 1$ (treated)

$$a_1 = \alpha_{11} + \gamma_1 \underbrace{\left(\sum_{i=1}^{n} \frac{T_i Y_i}{\hat{\pi}_i} \right)}_{\textbf{weight}} \qquad b_1 = \alpha_{10} + \gamma_1 \underbrace{\left(\sum_{i=1}^{n} \frac{T_i (1 - Y_i)}{\hat{\pi}_i} \right)}_{\textbf{weight}}$$

γ^* are renormalization terms

MODEL (T Binary Treatment, Y Binary Outcome, X Confounders, k Large)

Posterior Distribution: $p_T \mid \mathbf{Y}, \boldsymbol{\pi}(\mathbf{X}) \sim \text{Beta}\,(a_T, b_T)$

$$a_1 = \alpha_{11} + \gamma_1 \underbrace{\left(\sum_{i=1}^{n} \frac{T_i Y_i}{\hat{\pi}_i} \right)}_{\textbf{weight}} \qquad b_1 = \alpha_{10} + \gamma_1 \underbrace{\left(\sum_{i=1}^{n} \frac{T_i (1 - Y_i)}{\hat{\pi}_i} \right)}_{\textbf{weight}}$$

$$a_0 = \alpha_{00} + \gamma_0 \underbrace{\left(\sum_{i=1}^{n} \frac{(1 - T_i) Y_i}{1 - \hat{\pi}_i} \right)}_{\textbf{weight}} \qquad b_0 = \alpha_{01} + \gamma_0 \underbrace{\left(\sum_{i=1}^{n} \frac{(1 - T_i)(1 - Y_i)}{1 - \hat{\pi}_i} \right)}_{\textbf{weight}}$$

γ^* are renormalization terms

COMMENTS

Uses Bayesian computational procedures but is **not** Bayesian

- Incorporates **uncertainty** from propensity score estimation
- Maintains **separation** between treatment and outcome
- Outcome model does not assume a **parametric** function of treatment
- Simple diagnostic tools to assess **balancing properties**

Key Principles

- Reflect all uncertainty in estimates
- Adopt a design-based approach
- Avoid strong parametric specifications-likely in settings with many confounders to get model very wrong
- Adhere to causal inference assumptions

DOES UNCERTAINTY IN STEP 1 MATTER?

500 simulations, n = 1000, k= 100 (18 $\beta_j \neq 0$), $P(Y_i = 1) \approx 0.10$

$\hat{\Delta}$	Bias	95% CI Width	Coverage 95% CI
Integrated Propensity Score			
Student-$t_3(0, 2.5)$	-.011	.220	95.2%
Horseshoe Priors	.016	.110	93.0%
Bayesian Additive Regression Trees	.011	.123	96.8%
Mean Propensity Score			
Student-$t_3(0, 2.5)$	-.001	.095	**79.2%**
Horseshoe Priors	.018	.092	**86.0%**
BART	.015	.093	**87.2%**
Other Methods			
Naive Estimate	.030	.092	73.0%
IPW	-.001	.151	92.8%
TMLE	.006	.075	81.4%

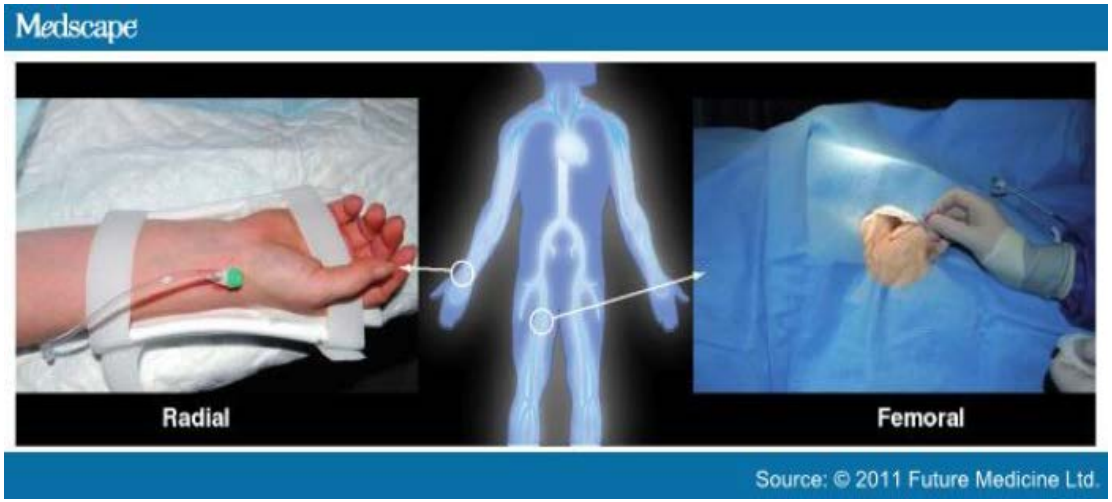
Bottom Line: useful to integrate over the propensity score distribution for large k

TWO EXAMPLES

1. General setting: radial or femoral artery access for PCI
2. High-dimensional setting: drug eluting versus bare metal coronary stents

RADIAL VS FEMORAL PCI

- Radial artery access permits easier access and easier closure
- Large number of patients undergoing both procedures
- Not particularly well studied and of growing importance in the U.S.
- Marked heterogeneity in predisposition to bleeding
- Significant treatment selection (healthier patients undergo transradial procedures)



Does **radial artery access** cause fewer complications compared to **femoral artery access**?

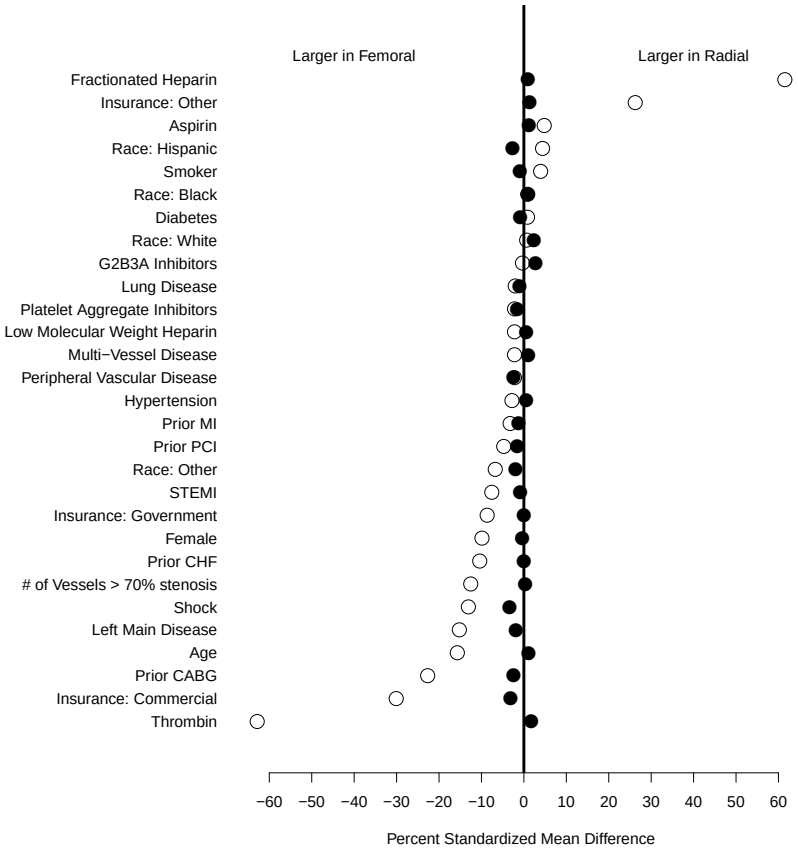
RADIAL VS FEMORAL ARTERY ACCESS

- **Registry** with data on **40,126 patients** PCI patient, 12.9% undergo PCI via radial artery (versus femoral artery)
- **Outcome:** in-hospital bleeding or vascular complications
- Unadjusted **Difference** (Radial - Femoral): $_{-2.30}$**-2.04%**$_{-1.80}$

	All Subjects	
	Intervention	
	Radial	Femoral
No. of Procedures	5192	35022
Mean Age [SD]	63 [12]	65 [12]
Female	25.3	29.8
Race		
White	89.6	89.4
Black	3.3	3.2
Hispanic	4.3	3.5
Asian	1.8	1.7
Native American	0.02	0.07
Other	1.0	2.2
Health Insurance		
Government	46.0	50.3
Commercial	4.8	13.4
Other	49.2	36.3
Comorbidities		
Diabetes	33.1	32.7
Prior CHF	9.4	12.7
Prior PCI	32.0	34.3
Prior myocardial infarction (MI)	28.7	30.1
Prior bypass surgery	8.4	15.7
Hypertension	79.6	80.7
Peripheral vascular disease	12.1	12.8
Smoker	24.8	23.1
Lung disease	13.7	14.4

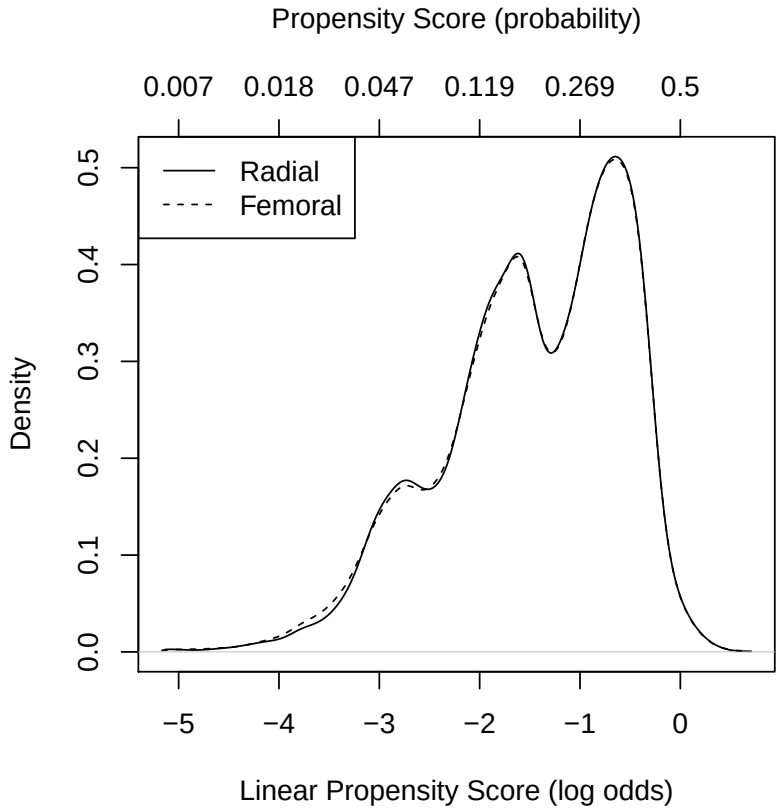
	All Subjects	
	Intervention	
	Radial	Femoral
No. of Procedures	5192	35022
Cardiac Presentation		
Multi-vessel Disease	10.3	10.9
Number of Vessels > 70% stenosis	1.49	1.58
Left main Disease	3.7	7.2
ST-elevated MI	38.9	42.6
Shock	0.44	1.8
Drugs Prior to Procedure		
Heparin (unfractionated)	87.3	61.7
Heparin (low weight molecular)	3.83	4.27
Thrombin	25.5	54.9
G2B3A inhibitors	26.7	26.8
Platelet Aggregate inhibitors	85.8	86.6
Intra-Aortic Balloon Pump	0.10	0.55
In-Hospital Complication, %	0.69	2.73
Mean Difference, % (95% CI)	-2.04 (-2.30, -1.80)	

ASSESSING VALIDITY OF POSITIVITY ASSUMPTION



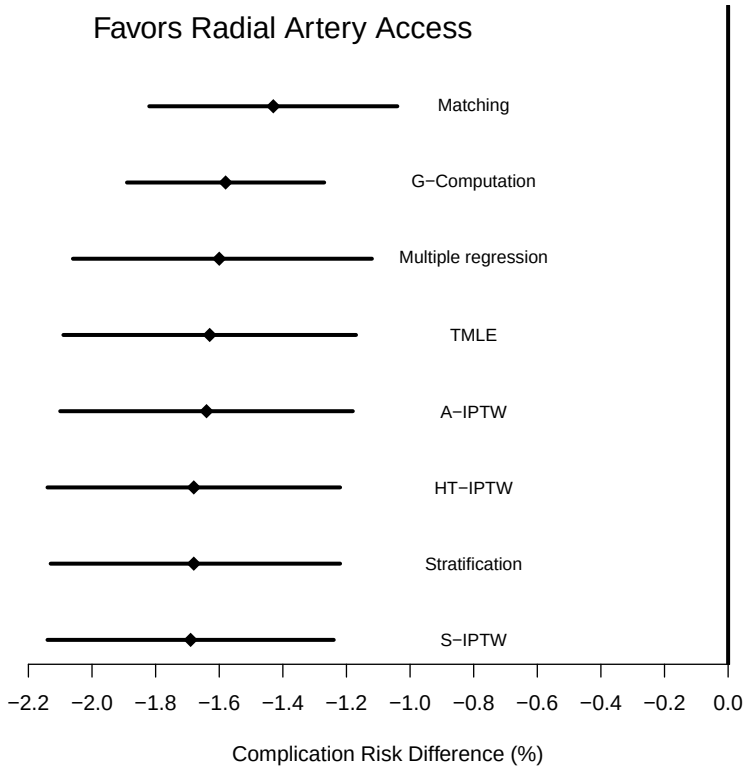
Mean(Radial)-Mean(Femoral)

Filled circles = Matched



Matched Subjects, $\hat{\Pi}_{T_i}$

RADIAL VS FEMORAL PCI



Assumptions

- **SUTVA**: practice makes perfect (physician random effect)
- **Ignorability**:
  - Omitted confounder: odds of radial $\geqslant 2.5\times$ femoral to change findings
- **Known Subgroups**:
  - Women: -2.67% (se = 0.43%)
  - Men: -1.00% (se = 0.58%)

$$\text{NNT} = {}_{46}\mathbf{49}_{51}$$

Source: Kunz, Rose, et al., in Methods in Comparative Effectiveness Research, Eds: C Gatonis, S. Morton, Chapman & Hall/CRC Biostatistics Series, 2017

DRUG ELUTING (DES)VERSUS BARE METAL (BMS) CORONARY STENTS*

- DES (approved 2003) and BMS (approved in 1990s) frequently implanted to keep treated arteries clear & supported
- DES improves target-vessel revascularization (TVR) more than BMS
- DES associated with late stent thrombosis (death)
- Have **9000** patients and **500** confounders

Do **DES** cause fewer revascularizations compared to **BMS**?

MASSACHUSETTS, 2011

	Stent Type	
Characteristic	BMS	DES
Outcomes, %		
1 Year Mortality	**10.2**	**3.3**
1 Year TVR	9.0	6.5
Confounders		
Age, yrs	66.4	63.7
STEMI, %	35.7	18.2
Cardiomyopathy or LVSD, %	11.1	8.4
Emergent, %	38.3	20.3
Shock, %	3.8	0.8

Source = Mass-DAC

Spertus and Normand, 2017 (under review)

DRUG ELUTING (DES)VERSUS BARE METAL (BMS) CORONARY STENTS

- Have **8718** patients and **495** confounders
- Clinical data collected prospectively using systemized tool, merged with willing data including all 5-digit present on admission ICD-9 codes occurring in > 10 patients
- **Confounders:** patient characteristics (age, sex, race, weight, etc.), morbidities (diabetes, heart failure, etc.), extent of blockage, ejection fraction, number of diseased vessels, etc., and billing codes.
- **Outcomes:** 1-year target vessel revascularization (PCI on same vessel or any CABG surgery)
- **Falsifiability outcome:** 30-day mortality
- **Treatment:** 65% DES vs 35% BMS

MODELS (T Binary Treatment, Y Binary Outcome, X Confounders)

Step 1: Treatment Model

$$T_i \sim \text{Bern}(\pi(\mathbf{X}_i))$$

$$\pi(\mathbf{X}_i) = \text{logit}^{-1}\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}\right)$$

$\pi(\mathbf{X}_i)$ = propensity score

- Priors for β_j: $t_3(0, 2.5)$ or Horseshoe
- Bayesian additive regression trees (BART)
- IPTW

Step 2: Outcome Model

$$Y_T \mid n_T, p_T, \pi(\mathbf{X}) \sim \text{Binomial}(n_T, p_T)$$

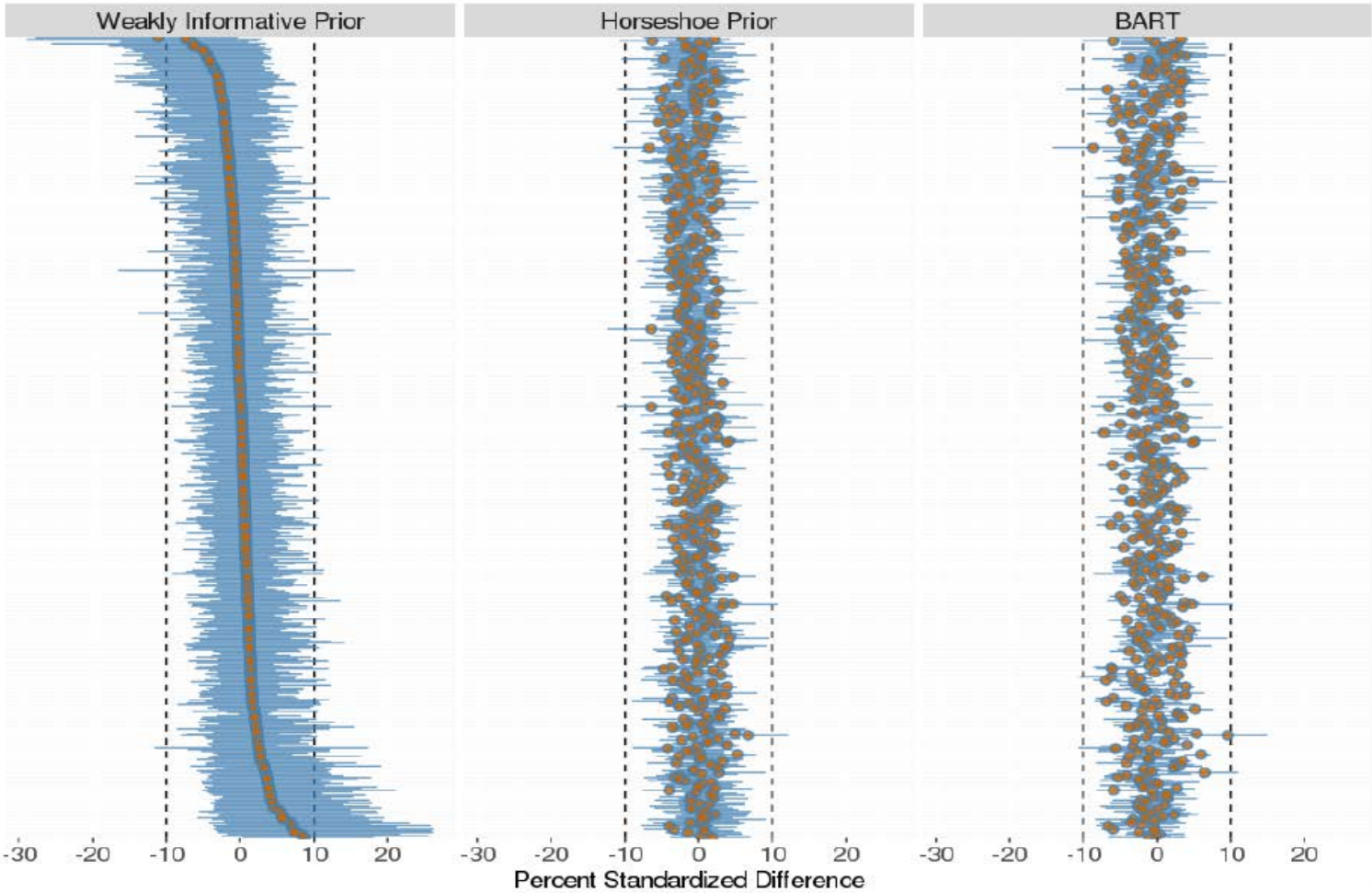
$$p_T \mid \alpha_{T0}, \alpha_{T1} \sim \text{Beta}(\alpha_{T0}, \alpha_{T1})$$

n_T = number of subjects receiving treatment T
p_T = probability of outcome under treatment T
Independent samples in treatment groups

- *A-posteriori*

$$p_T \mid \mathbf{Y}, \pi(\mathbf{X}) \sim \text{Beta}(a_T, b_T)$$

WEIGHTED STANDARDIZED MEAN DIFFERENCES IN CONFOUNDERS



CAUSAL ESTIMATES

		Mortality		Weight
Approach	**% 1-Year TVR**	**% 30-Day**	**% 1-Year**	**Mean (Max)**
Naive	-2.4 (-3.6,-1.3)	-3.4 (-4.1, -2.6)	-6.9 (-8.0, -2.6)	NA

CAUSAL ESTIMATES

		Mortality		Weight
Approach	**% 1-Year TVR**	**% 30-Day**	**% 1-Year**	**Mean (Max)**
Naive	-2.4 (-3.6,-1.3)	-3.4 (-4.1, -2.6)	-6.9 (-8.0, -2.6)	NA
IPTW	-3.3 (-5.5, -1.1)			2.0 (54.9)
BART	-4.1 (-5.7,-2.6)			2.0 (28.2)
Horseshoe	-3.9 (-5.2, -2.6)			2.0 (29.6)
$t_3(0, 2.5)$	-3.3 (-5.5, -1.1)			2.3 (100.6)

CAUSAL ESTIMATES

		Mortality		Weight
Approach	**% 1-Year TVR**	**% 30-Day**	**% 1-Year**	**Mean (Max)**
Naive	-2.4 (-3.6,-1.3)	-3.4 (-4.1, -2.6)	-6.9 (-8.0, -2.6)	NA
IPTW	-3.3 (-5.5, -1.1)	-1.0 (-2.0, 1.8)	-2.9 (-4.9, -0.3)	2.0 (54.9)
BART	-4.1 (-5.7,-2.6)	**-1.3 (-2.1, -0.6)**	-3.3 (-4.5, -2.1)	2.0 (28.2)
Horseshoe	-3.9 (-5.2, -2.6)	**-1.2 (-2.0, -0.5)**	-3.2 (-4.5, -2.0)	2.0 (29.6)
$t_3(0, 2.5)$	-3.3 (-5.5, -1.1)	-0.3 (-1.7, 2.5)	-2.1 (-4.2, 0.9)	2.3 (100.6)

OUTLINE

- Introduction
  - Three examples
  - What is CER and why important?
  - Main characteristics of CER
- Methodology
  - Assumptions and causal parameters
  - Approaches
    - Typical setting
    - High-dimensional data
  - Examples
- Closing Remarks
  - Research needs

METHODOLOGICAL NEEDS FOR CER

1. Pragmatic trials
2. Missing data
3. Non-constant treatment effect

PRAGMATIC TRIALS

Compare $\geqslant$ **2** medical interventions **directly relevant** to clinical care.

- Features:
  - **Broad** eligibility criteria
  - Medical management consistent with **usual care** (no blinding, no extra office visits)
  - Measure all outcomes **important** to patients (patient self-assessments)
- Problems:
  - High rates of non-adherence can **bias** non-inferiority/equivalence trial **towards equivalence**
  - High rates of **loss to follow-up** problematic
  - Require **larger** sample sizes because:
    - Comparing two active treatments (vs active and placebo)
    - Small treatment effects in **heterogeneous populations**
  - **Bias**: unblinded treatment assignments, unblinded clinical assessments, and patient self-assessments

MITIGATING LIMITATIONS OF PRAGMATIC TRIALS

- Design Strategies:
  - Include objective (survival, test results) in addition to subjective (patient self-assessments)
  - Baseline measure of self-assessments
- Analytical Strategies
  - ITT **tends towards** making the two drugs similar (D'Agostino, Massaro, Sullivan; 2003)
    - Use of instrumental variables to estimate the ITT effect and the Per Protocol effect
  - Measurement error in self-assessments

MISSING DATA

(T Treatment, Y Outcomes, $\Pi_T = P(T \mid \mathbf{X})$)

Let $R = 1$ if outcome is observed and 0 if missing & assume **ignorable** missingness:

$$\begin{aligned} P(Y, R, T, \mathbf{X}) &= P(Y \mid R = 1, T, \mathbf{X}) \times P(R \mid T, \mathbf{X}) \times P(T \mid \mathbf{X}) \times P(\mathbf{X}) \\ &= Q_Y \times \Lambda_R \times \Pi_T \times Q_X \end{aligned}$$

1. Treatment effect still depends **only** on Q_Y **and** Q_X: $E_X\left(E(Y \mid T = 1, \mathbf{X})\right) - E_X\left(E(Y \mid T = 0, \mathbf{X})\right)$
2. Π_T is the **propensity score** (nuisance): $\Pi_T = P(T = 1 \mid \mathbf{X})$
3. Λ_R describes **missing mechanism** (nuisance): $\Lambda_R = P(R \mid T, \mathbf{X})$

ESTIMATORS

(T Treatment, Y Outcomes, R Missingness, $\Pi_T = P(T \mid \mathbf{X})$)

Can be used in conjunction with the estimators described previously:

- Horvitz-Thompson (1952):

$$\begin{aligned} \text{IPTW}_{\text{HT}} &= \frac{1}{N}\sum_{i=1}^{N}\frac{T_i Y_i}{\Lambda_{R_i}\Pi_{T_i}} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-T_i)Y_i}{\Lambda_{R_i}(1-\Pi_{T_i})} \\ \Lambda_R &= P(R \mid T, \mathbf{X}) \end{aligned}$$

Need **more empirical** studies of these approaches

NON-CONSTANT TREATMENT EFFECTS

1. **Unknown** which subgroups and **have** all measured confounders
   - Stratification on propensity score (weak)
   - Extend approach of Zigler and Dominici (2014)
2. **Unknown** which subgroups and **do not have** all measured confounders
   - If heterogeneity present, then different instruments will lead to different treatment effect estimates (seemingly contradictory results)
   - See **A Basu** (Statistics in Biosciences CER issue) when unmeasured confounders **moderate** treatment effects

WHAT WAS NOT DISCUSSED TODAY

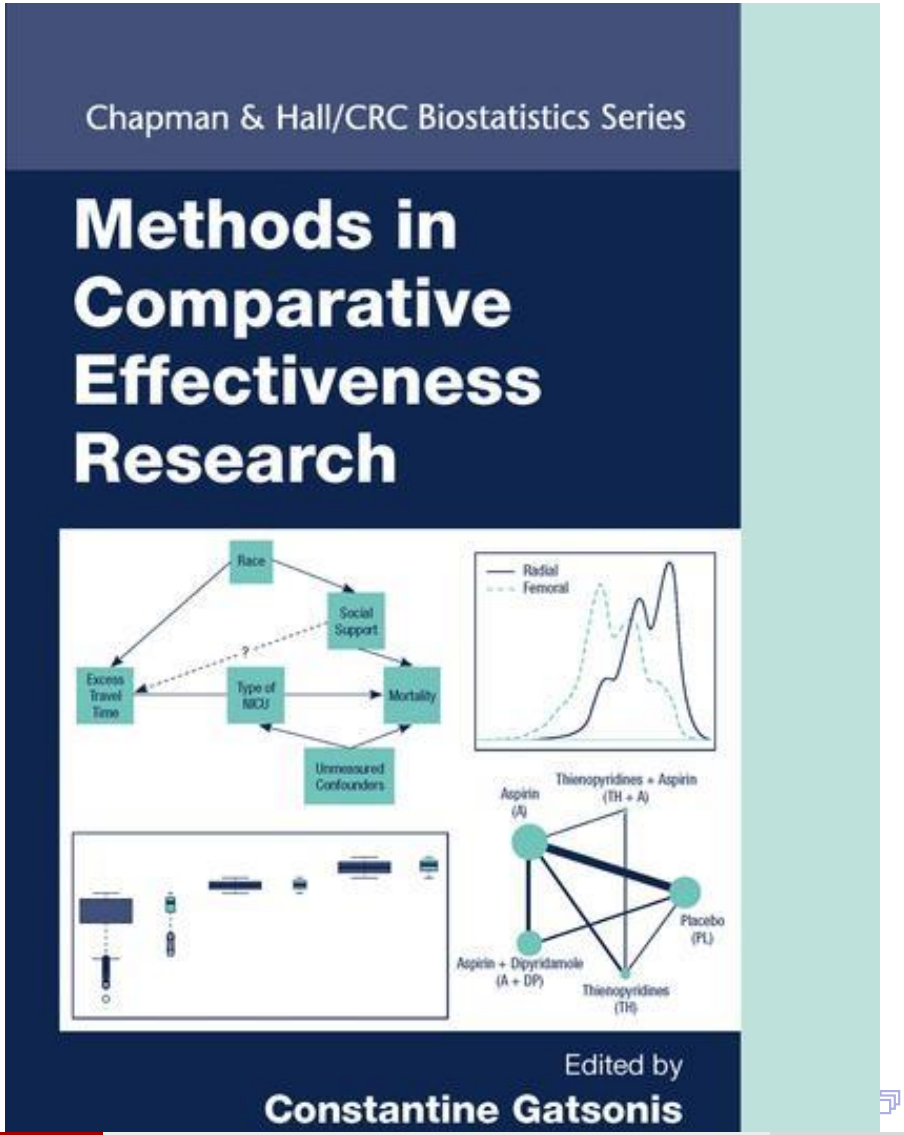
- Approaches when ignorability of treatment assignment is unlikely
  - Instrumental variables
  - Bounding inferences using external information about size of residual confounding
- Exploratory and confirmatory approaches to non-constant treatment effects
- Average effect of treatment on the treated
- Mathematical modeling
  - Decision trees, Markov models, individual microsimulation, dynamic transmission models, discrete event, agent-based simulation

THANK YOU

PAPERS

- Statistics in the Biosciences 2011; 3(1) Issue.
- Statistics in Medicine 2010; 29(19) Issue.
- Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. Ann Intern Med 2009;151:206-9.
- Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research. National Academies 2009.
- Institute of Medicine. Learning What Works Best: The Nations Need for Evidence on Comparative Effectiveness in Health Care. National Academies. 2008

RELEASED FEB. 2017



REFERENCES

- Basu A. Estimating decision-relevant comparative effects using instrumental variables. Statistics in the Biosciences, 2011; 3(1):6-27.
- Carvalho CM, Polson NG., Scott JG. Handling sparsity via the horseshoe. Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, 2009; 5:73–80.
- D'Agostino Sr. RB, Massaro JM, Sullivan LM. Non-inferiority trials: design, concepts and issues - the encounters of academic consultants in statistics. Statistics in Medicine, 2003; 22:169-186.
- Gelman A, Jakulin A, Grazia P, Su YS. A weakly informative default prior distribution for logistic and other regresssion models. Annals of Applied Statistics, 2008; 2(4):536-549.
- Gruber S, van der Laan M. An application of targeted maximum likelihood estimation to the meta-analysis of safety data. Biometrics, 2013; 69:254-262.
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 1952; 47:663-685.

REFERENCES (CONT.)

- Hoeting JA, Madigan D, Raftery AE, Vollnksy CT. Bayesian model averaging: a webinar. Statistical Science, 1999; 14(4):382-401.
- Kaplan D, Chen J. A two-step Bayesian approach for propensity score analysis: simulations and case study. Pyschometrika, 2012; 77:581–609.
- Kunz L, Rose S, Spiegelman D., Normand S-L. An overview of statistical approaches for comparative effectiveness research. Chapter in Methods in Comparative Effectiveness Research. Eds: Gatsonis C. and Morton S. CRC Press, 2017.
- Ravikumar P, Lafferty J, Liu H, Wasserman L. Sparse additive models. Journal of the Royal Statistical Society, Series B 2009; 71:1009-1030.
- Ridsgeway G, McCaffrey D, Morral A, Burgette L, Griffin BA. Toolkit for weighting and analysis of nonequivalent groups: A webinar for the twang package. 2014, R Vignette, RAND.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 1984; 79:516-524.

REFERENCES (CONT.)

- Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association, 1994; 89:846-866.
- Rubin DB. Bayesian inference for causal effects:the role of randomization. Annals of Statistics, 1978; 6(1):34–58.
- Saarela O, Stephens DA, Moodie EEM, Klein MB. On Bayesian estimation of marginal structural models. Biometrics, 2016; 71:279–301.
- Schneeweis S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology, 2009; 20, 512.
- Snowden KM, Rose S, Mortimer KM. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. American Journal of Epidemiology, 2011; 173(7):731-738.
- van der Laan M, Rose S. Targeted Learning: Prediction and Causal Inference for Observational and Experimental Data. New York: Springer.

REFERENCES (CONT.)

- Zigler CM, Dominici F. Uncertainty in propensity score estimation. Bayesian methods for variable selection and model-averaged causal effects. Journal of the American Statistical Association, 2014; 109:95-107.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society B, 2005; 67(2):301-320.