An Overview of Causal Inference and its Applications in Clinical Trials

Thamban Valappil, Ph.D.
Division of Biometrics IV, CDER/FDA

BASS XIV, 2007

Acknowledgments:

Mark van der Laan, Ph.D.
Romain Neugebauer, Ph.D.
Kelly Moore

Disclaimer

Views expressed in this presentation are of those of the speaker and not necessarily of the Food and Drug Administration.

Outline

- Role of Causal Inference
- Randomization
- Bias of standard methods
- Noncompliance issues
- Point Treatment vs. Time Varying Treatment Problems
- Marginal Structural Models
- Summary

Role of Causal Inference

- Clear framework for explaining the role of randomization, confounding etc..
- Estimate the causal effect of a treatment
- Set bounds on causal effect
- Control confounding or effect modifiers

What is Causal Effect ?

- The causal effect in an individual is the difference between the potential outcomes due to different treatments under study.

- Potential outcomes are the values that would have been observed had each treatment being given to a patient.

- Causal Effect is defined in terms of a comparison of the counterfactual failure times associated with different exposure/treatment under study. For a typical randomized study, we compare the average effect of the treatment in the population.

What is Causal Effect ?

- The goal of many clinical trials or epidemiologic studies is to quantify the causal effect of a treatment (exposure) on an outcome.

- In contrast, the most commonly used statistical methods provide measures of association, not necessarily of causal effect.

- These association measures may lack a causal interpretation even when the investigator "adjusts for" all potential confounders in the analysis of a properly designed study.

Randomization

- It produces groups that are similar with respect to their baseline characteristics and prognostic factors.
- Balances measured and unmeasured confounders in the treatment arms.
- Provides a valid basis of inference (enables probability statements or p-values).
- Randomization allows us to make inference regarding cause and effect of a treatment and a clinical outcome.

What's wrong with observational studies?

- Lacks randomization and the estimates for the treatment effect may be seriously biased.
- If the treatments are (Drug A and Drug B), there is no guarantee that study populations (those getting Drug A and Drug B) will be equivalent with respect to the risk of the disease.
- The ideal situation would have been to give everyone Drug A, record their response, go back in time and then give Drug B.

Bias of Standard Methods

Consider an observational study and the objective is to estimate the change in hazard due to exposure.

The survival model is,

T = f (E, C), where E : treatment exposure

C : Confounder

- Usual approach to the estimation of the effect of a time-varying treatment on survival has been to model the hazard of failure at time t as a function of past treatment history using time-dependent Cox proportional hazard or other regression models.

Intent-To-Treat (ITT) analysis

- Preserves the baseline comparability between treatment groups as it retains all randomized subjects.
- It does not exclude noncompliers or drop-outs from the analysis.
- Comparison of treatments is based on the difference of the average response in the treatment groups.

Noncompliance in Clinical Trials

- Non-informative:

  The risk of discontinuation or noncompliance is independent of the outcome. i.e. Compliant and noncompliant subjects would have identical risk.

  This is a strong assumption and is seldom true in most of the clinical trials.

- Informative:

  Discontinuation or noncompliance provide information and influences the outcome.

Noncompliance in Clinical Trials

Complier analysis:

- Compare only those subjects who fully complied with their assigned treatment and exclude noncompliers from the analysis.
- Compliance or noncompliance occurs after randomization and excluding them can bias the treatment evaluation.
- The prognostic effect of noncompliance cannot be separated from the actual treatment effect.
- The assumption that would justify this analysis is “ignorable non-compliance” (Rosenbaum and Rubin- *Biometrika* 1983; 70:41-55)
- Intent-to-Treat analysis should be the preferred primary analysis.

Example-1: Coronary Drug Project

A randomized, multi-center, double-blind, placebo-controlled trial comparing clofibrate (1.8 g per day) to placebo for the treatment coronary heart disease. (NEJM, 1980 (303), 1038-1041).

Objective: The Coronary Project was carried out to evaluate the efficacy and safety of several lipid influencing drugs in the long term treatment of Coronary Heart Disease (CHD)

Enrollment and Patient Characteristics

- Enrollment Period: March 1966 to October 1969, 53 clinical centers
- Men
- Age from 30 to 64 years
- To qualify, electrocardiographic evidence of MI occurred not less than 3 months previously
- Clinical visits occurred every four months for a minimum of 5 and maximum of 8.5 years

Example-1: Coronary Drug Project

Treatment	5-year Mortality Rate (# of patients)	P-Value
Clofibrate Placebo	20.0% (n=1103) 20.9% (n=2789)	P-value= 0.55

No significant difference between Clofibrate and Placebo (p=0.55).

Treatment	5-Year Mortality rate (# of patients)
Clofibrate	
Compliance: ≥80%	15.0% (n=708)
<80%	24.6% (n=357)
	(p-value=.00011)
Placebo	
Compliance: ≥80%	15.1% (n=1813)
<80%	28.2% (n=882)
	P-value=4.7x E-16

Compliance and Mortality

- Even if there is a decrease in mortality between the good and poor compliance to clofibrate, this effect could be confounded by the major differences in the patients characteristics and prognosis in the subgroups.

- This subgroup analysis can seriously bias the results since the prognostic effect of non-compliance and treatment cannot be delineated.

- Other Subgroup analysis:

They have compared compliant (≥80%) clofibrate subjects (15%) to all placebo subjects (19.4%) –shows **significance**! So what?

Issues with the study

- Based on the ITT analysis, clofibrate was not more beneficial than placebo
- Based on the post-hoc explanatory analysis, the decrease in mortality of subjects complying with clofibrate shouldn't be attributed to clofibrate.
- Subgroup analysis of patients based on characteristics measured post-randomization will not provide a scientifically valid comparison.
- Subgroup analyses are unreliable and can introduce serious bias when patients are selected into groups that are good or bad with respect to compliance or response.

What is the role of Causal Inference in addressing these issues?

Potential Outcomes

The data show all potential outcomes under two different treatments. Y(0) and Y(1) denotes the number of years lived using Drug A and Drug B.

Y(0)	Y(1)
10	13
4	2
3	1
4	2
6	2
5	4
7	6
6	3

Observed Outcomes

Y(0)	Y(1)
10	?
?	2
?	1
4	?
6	?
?	4
7	?
?	3

Counterfactuals

- Counterfactuals represent the outcomes under circumstances that may not have actually occurred

Binary Treatment : $A = \begin{cases} 1 \text{ if treated} \\ 0, \text{ otherwise} \end{cases}$

Y: Observed Outcome

Y(0) = potential counterfactual, if untreated
Y(1) = counterfactual, if treated

Using a consistency relation, $Y = Y(1)\,A + Y(0)(1-A)$

Association and Causal Effect

Association:

$$\alpha = E(Y|A=0) - E(Y|A=1)$$

If $\alpha \neq 0$, does that mean treatment A has causal effect on Y ?

In a randomized study : Yes

Observational study : Not necessarily (due to confounding)

Causal Effect:

$$\theta = E\{Y(1)\} - E\{Y(0)\}$$

In general,

- $\alpha \neq \theta$
- α is identifiable because it depends only on (A, Y)
- θ is ***not*** easily identifiable because it depends only on counterfactuals

Association and Causal Effect

But, if the treatment is randomized then

Y(0), Y(1) independent of treatment A. I.e. Treatment assignment is independent of the potential outcomes.

Therefore,

$\theta = E\{Y(1)|A=1\} - E\{Y(0)|A=0\}$

$= E(Y|\ A=1) - E(Y|\ A=0)$

$= \alpha$

So, randomization:

- makes θ and α equal
- makes θ identifiable, since α is a function of (A,Y)

Association and Causal Effect

Suppose Y(0), Y(1) are not independent of Treatment A

Then $\theta \neq \alpha$

But, we can find a covariate(s) L such that Y(0) and Y(1) are independent of A| L , then θ is identifiable.

When would that be true ?

In a blocked experiment where Treatment A is randomized within levels of L

In an observational study L represents all important confounders. The best we can hope is that the above condition holds approximately.

Consistency and Exchangeability

- In randomized experiments association *is* causation because
  - experimental treatment assignment produces consistency
  - randomization produces exchangeability
- In observational studies
  - If no consistency, then counterfactuals are not well defined
  - If counterfactuals are not well defined then causal effects are not well defined

Consistency and Exchangeability

- In a randomized study, the outcomes of the patients who received treatment and control are representative samples of the outcomes in the population

- Therefore, the average outcome of those randomized to receive treatment is an unbiased estimator for $E(Y_1)$

- Likewise, the average outcome of those randomized to receive control is an unbiased estimator for $E(Y_0)$

- Therefore, the difference of these sample average is an unbiased estimate of the population average causal effect,

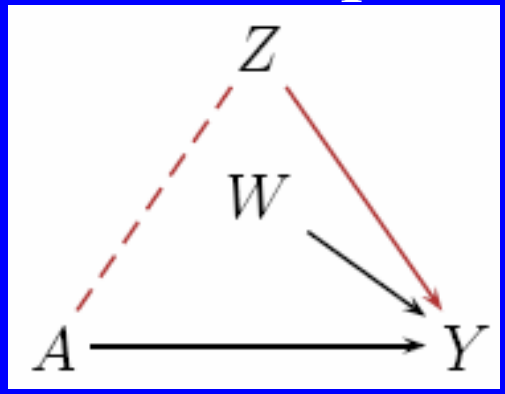
$$E(Y_1) - E(Y_0)$$

Confounding in randomized studies

Randomized study

What if the subjects do not comply with their assigned treatment and there are informative drop-outs?

Like observational studies, randomized trials can also be subject to empirical confounding and confounding due to informative dropouts and can bias the estimates.



Exchangeability and confounding

Unadjusted Analysis:

The regression of the outcome on the treatment relies on the assumption of exchangeability,

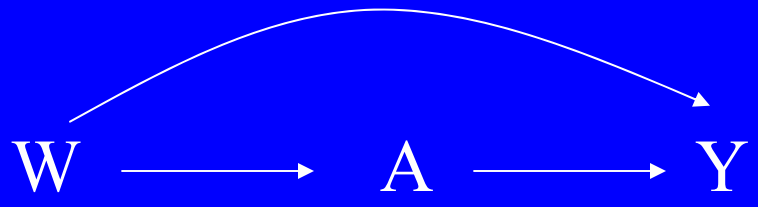
$$E(Y \mid A=a) = E\ (Y_a)$$

- The outcomes collected on the treated can be used to infer about the outcomes we could have seen on the untreated had the untreated been treated (and vice versa).
- The counterfactual risk in the treated equals the counterfactual risk in the untreated

When will the assumption of exchangeability of the treated and non-treated be violated?

Exchangeability and confounding

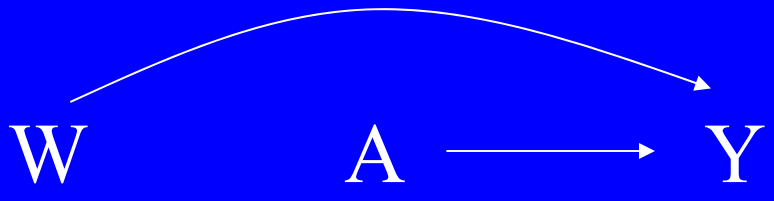
When there is confounding ,i.e., when a variable (collected or not) affects both the treatment and outcome:



The unadjusted analysis allows investigation of the marginal effect of A on Y only in the absence of confounding.

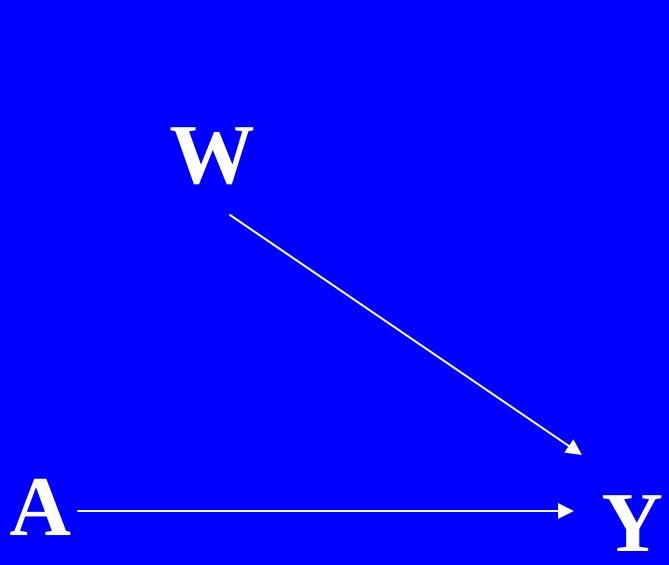
Exchangeability and confounding

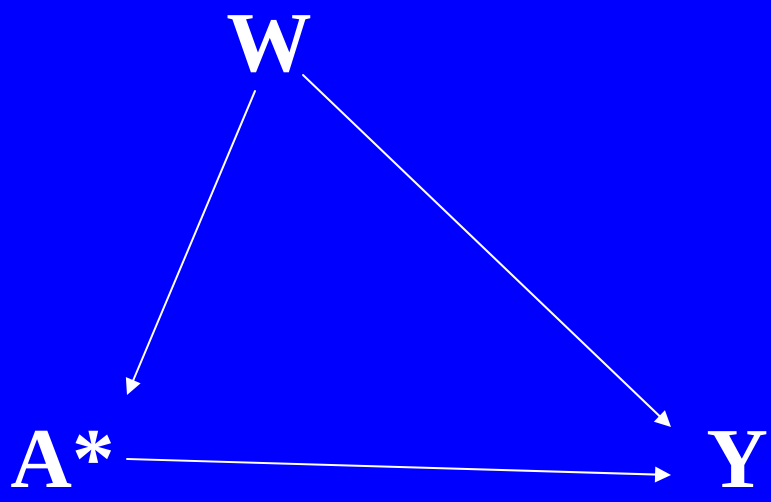
- Not realistic in most observational studies
- Acceptable in most randomized studies where the treatment is assigned completely at random, i.e. independently of the patients' characteristics:



g(A | subject's characteristics) = g(A)

How about Randomized study ?





Non-compliance (A*) can lead to confounding

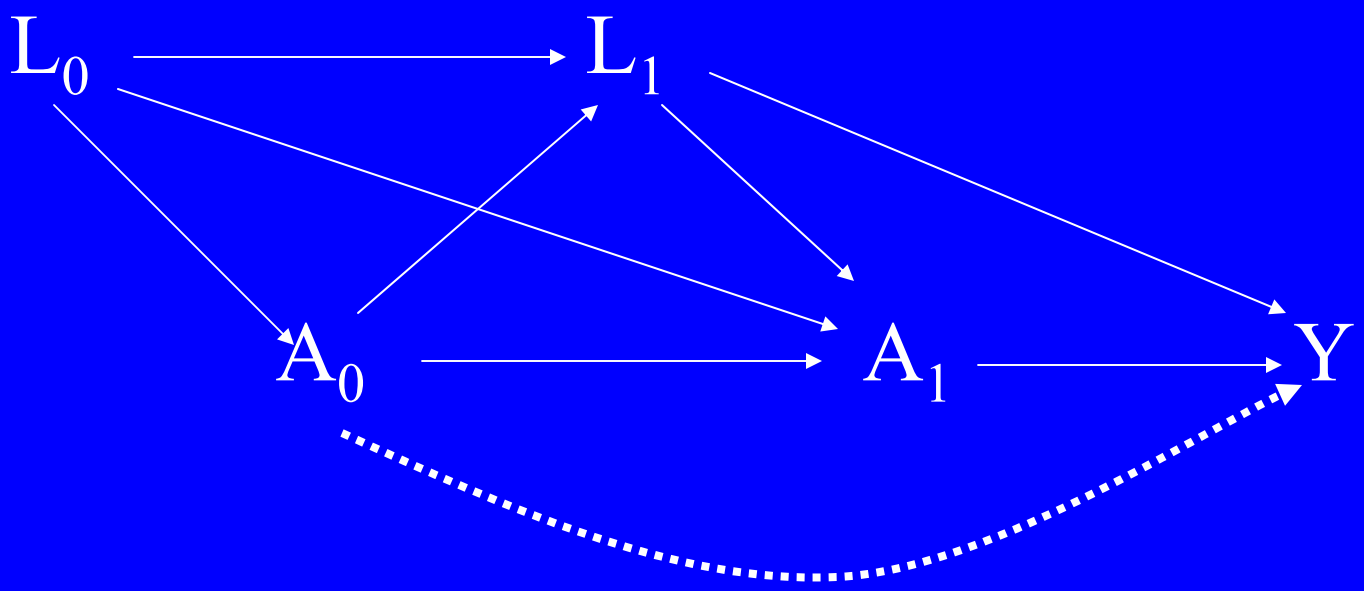
Simplest causal inference problem:

- What is the population-level effect of a treatment, A, on a clinical outcome, Y ?

More complex causal inference problem:

- What is the population-level effect of a treatment experienced over time, A(0), . . . ,A(K), on a clinical outcome, Y ?

Typical Causal Problem



Notation:

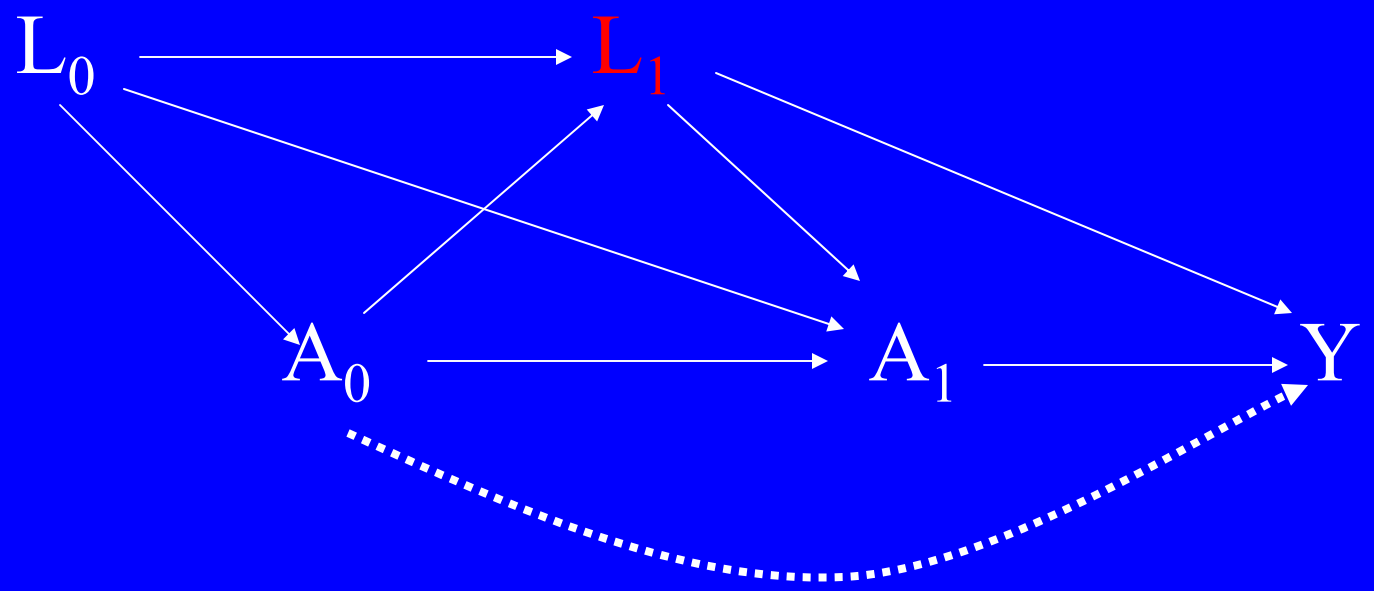
Two time points and variables at each time point are dichotomous

L_0, L_1 = Indicator that a confounding variable is present at time 0 and 1

A_0, A_1 = Indicator for treatment observed at time 0 and 1

Y = Outcome

Typical Causal Problem



For multiple time points, IPTW method could provide unbiased estimates of the casual parameters (Robins et. al).

Defining Causal Parameters

Question of interest:

What is the causal effect of A on Y?

- To answer this questions, one collects n independent and identically distributed (i.i.d.) observations of O.
- $X = (W, Y_0, Y_1)$
- Using the counterfactual framework, we can now formally represent the concept of causal effects with causal parameters.

Defining Causal Parameters

Causal effect representation:

- marginal effect:
  - individual level: $Y_1 - Y_0$
  - population level:

  $\beta = E(Y_1) - E(Y_0)$ or $\beta = E(Y_1)/E(Y_0)$

- Conditional effect
  - individual level: $Y_1 - Y_0$
  - population level: $\beta(V) = E(Y_1 \mid V) - E(Y_0 \mid V)$

    where $V \subset W$

Point-Exposure vs. Time-varying Exposure Studies

- In a point exposure study, one usually models the probability of disease at one point in time as a function of exposure and pre-treatment (baseline) covariates.
- Time-dependent treatment studies are those where the effect of interest is a time-varying treatment or exposure.
- In a time-varying exposure study, the traditional or standard methods may be biased if time varying covariates are simultaneously confounders and intermediates. i.e. covariates are predictors of outcome and also predict subsequent exposure, and past exposure history predicts resulting covariate level (Robins 1991)

Counterfactuals - Recap

- If Y is the outcome, A is the treatment of interest, then the ideal scenario is where one observes, for each subject, Y_a, for each treatment level A=a.

- For example, if there is simply two levels of exposure (eg, treatment A= (1 , 0) where 1= yes; 0 = no), then each subject has in theory two counterfactuals, Y_0 and Y_1.

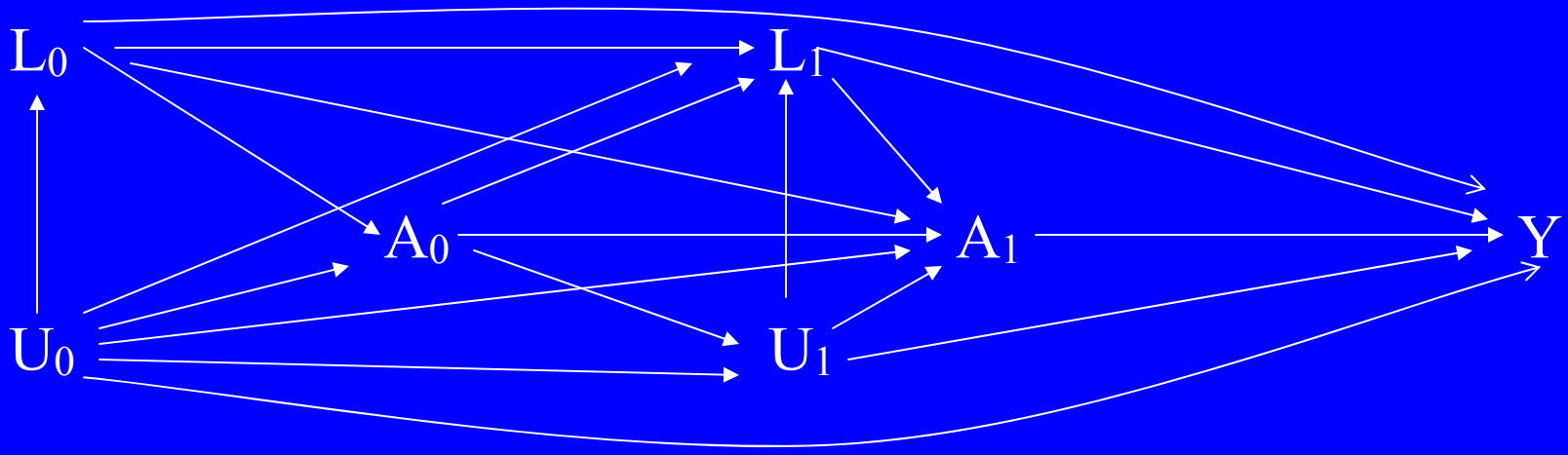
- To estimate specific causal effects, we then define parameters that relate, for instance, how the means of these counterfactuals differ as one changes a.

Challenges in Longitudinal Studies

- Now let us see how the history of treatment (or exposure) is causally associated with a future outcome in longitudinal studies.

- For example, let $L_{j,\, j=0,1}$ be the confounders, $A_{j,\, j=0,1}$ be the treatment/exposures and Y the outcome of interest.

- Assume that the association of interest is on the sum of treatment ($\Sigma A_j = A_0 + A_1 = a$) and the outcome Y, at the end of the study.



In standard analysis, one would adjust for the confounders to estimate the coefficient β in the following model,

$$E[Y \mid A_1 + A_0 = a, L_0 = l_0, L_1 = l_1] = \beta_0 + \beta_1 a + \beta_2 l_0 + \beta_3 l_1$$

This will not work, since L_1 is both a confounder of A_1, and is also on the causal pathway of A_0 and Y and does not have causal interpretation.

Simple Point Exposure or Treatment case

- That is, when one has an exposure or treatment of interest measured at one time point, and we want to find the causal association of the factor of interest on an outcome.

  Next few slides will briefly introduce few estimators including inverse probability treatment weighting, double robust, G-computation and targeted MLE.

Marginal Structural Models

Marginal structural models (MSMs) are causal models for the estimation of the causal effect of a time-dependent exposure in the presence of time-dependent covariates that may be simultaneously confounders and intermediate variables (Robins et. al 1998, 2000; Hernan et. al 2001).

Unlike conventional models, they represent causal effects based on the concept of potential outcomes. MSM estimation is a missing data problem.

Statistical framework for point treatment studies

- Observed data: $O = (W, A, Y)$
  - Baseline covariate: W
  - Treatment(s): A = (0,1)
  - Outcome: Y
- Full data: $X = (W, (Y_a)_{a \in A}) \sim F_X$, where F_X denotes the full data distribution
- Assumptions:
  - existence of counterfactuals
  - time-ordering assumption
  - consistency assumption: $Y = Y_A$, i.e. the observed outcome corresponds to one potential outcome.

Total Effects in Point Treatment Studies

Parameters of the distribution of the counterfactuals, Y_a

Examples (for binary treatment A (0,1):

marginal effect:

- individual level: $Y_1 - Y_0$
- population level: $\beta = E(Y_1) - E(Y_0)$ or
  $\beta = E(Y_1)/E(Y_0)$

Can be expressed as causal OR:

$$E[Y_1](1-E[Y_0])/\{E[Y_0](1-E[Y_1]\}=$$
$$P[Y_1=1](1-P[Y_0=1])/\{P[Y_0=1](1-P[Y_1=1]\}$$

(Expressed in terms of potential outcomes with and without exposure to the treatment)

Assumptions

1. Consistency Assumption: observed data, O is $O=(A,X_A)$ – i.e., the data for a subject is simply one of the counterfactual outcomes from the full data.
2. Randomization Assumption: $A \perp Y_a \mid W, \forall a$

   No unmeasured confounders for treatment. In other words: within strata of W, A is randomized
3. Experimental Treatment Assignment:

   All treatments are possible for all members of the target population,

   $P(A = a \mid W) > 0, \qquad \forall$ W.

Point Treatment case: Likelihood of Data

- The likelihood of the data factorizes into the distribution of interest and the treatment assignment distribution.

$$L(O) = P(Y \mid A, W)P(A \mid W)$$

MSM Estimation

- Traditional Unadjusted Analysis
- G-computation estimation: model for $E(Y \mid A,W)$
- Inverse weighting estimation: model for $g(A \mid W)$
- Double robust estimation: model for $E(Y \mid A,W)$ and $g(A \mid W)$
- Target maximum likelihood estimation: model for both $E(Y \mid A,W)$ and $g(A \mid W)$

MSM estimation

We will discuss three estimators:

- G-computation
- Inverse Probability of Treatment Weighted (IPTW)
- Double Robust (DR)

All assume no unmeasured confounding.

All account differently for confounding by modeling different nuisance parameters:

G-Computation algorithm

G-computation estimation relies on two assumptions:

- Randomization assumption (or no unmeasured confounders): $A \perp Y_a \mid W, \forall a$ which insures conditional exchangeability.
- Consistent estimation of the nuisance parameter

$$Q(a,W) = E(Y \mid A,W)$$

G-computation Method (Robins et. al)

- Based on the earlier assumptions,
  $E(Y|A=a,W) = E(Y_a|W)$
- Then, $E[Y_a] = \int_w E[Y_a \mid W = w] dP(w)$
- Which leads to G-comp. estimate of the counterfactual mean.

$$\hat{E}[Y_a] = \sum_{i=1}^{n} \frac{1}{n} \hat{E}[Y \mid A = a, W = W_i]$$

- Regress $\hat{E}[Y_a]$ vs. a to get an estimate of MSM.

G-computation implementation

The G-computation estimator of the MSM parameter can be implemented in a two-step procedure:

- estimate $Q(A,W) = E(Y \mid A,W)$ (e.g. parametrically after model selection). We denote the estimate with $Q_n(A,W)$
- regress $Q_n(a,W)$ on a and V according to the MSM, $m(a, V \mid \beta)$ (e.g. $\beta_0 + \beta_1 a$) for every possible a.

$$\beta_n = \arg\min_{\beta} \sum_i \sum_a (Q_n(a, W_i) - m(a, V_i \mid \beta))^2$$

G computation estimator correspond to the Maximum Likelihood Estimator

IPTW analysis

WHY?

The motivation behind the IPTW estimator is that weighting observations by their respective propensity score creates a pseudo-population in which treatment assignment is no longer confounded (Robins 1998).

- IPTW calculates the probability of an individual receiving the treatment they actually received conditional on their observed covariates. i.e. IPTW models treatment assignment to adjust for confounding and uses these as weights in regression.

- Define $g(a|W) = P(A=a|\ W)$.

- IPTW estimator is an extension to the longitudinal causal inference models of estimators proposed by Hortivtz &Thompson, Greenland, Rosenbaum, Robins & Rotnitzky and others…

IPTW Estimators

- Unlike in the observed data, the groups of treated and non-treated are exchangeable in the weighted data.
- Weighting removes confounding by creating a ghost data set where the treatment is randomized.
- General estimating function is

$$\frac{h(A,V)}{g(A \mid W)}(Y - m(A,V \mid \beta))$$

(for stratified by V⊂W MSM)

Example 1: (Hernan et. al, stats in medicine, 2002)

	L=0 (no confounder)			**L=1 (Confounder)**	
	A=0	**A=1**		**A=0**	**A=1**
1	Y_1	?	**5**	?	Y_5
2	Y_2	?	**6**	?	Y_6
3	?	Y_3	**7**	Y_7	?
4	Y_4	?	**8**	?	Y_8

	L	A	P(A=a|L)
1	0	0	3/4
2	0	0	3/4
3	0	1	1/4
4	0	0	3/4
5	1	1	3/4
6	1	1	3/4
7	1	0	1/4
8	1	1	3/4

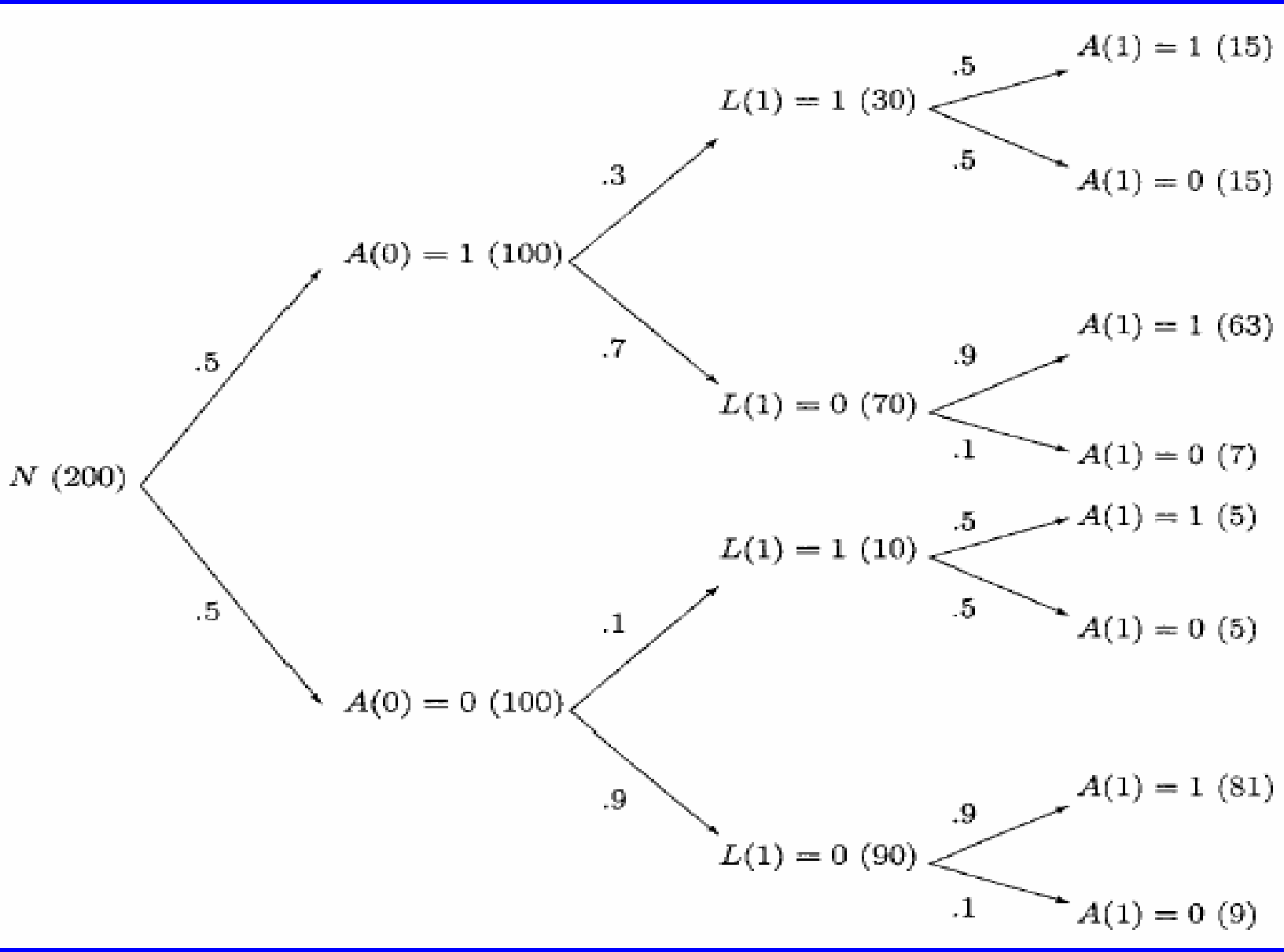
	L	A	$[P(A=a\|L)]^{-1}$
1	0	0	4/3
2	0	0	4/3
3	0	1	4
4	0	0	4/3
5	1	1	4/3
6	1	1	4/3
7	1	0	4
8	1	1	4/3

The inverse of the weights would be used in the IPTW method

Example: 2 (Hernan et. al, stats in medicine, 2002)

Consider a sequentially randomized trial with two treatments Zidovudine (A(1)=1) and placebo, (A(0)=1), where treatment at time k is randomly assigned with the randomization probabilities possibly depending upon covariate history and past treatment $A(k-1)$.

- Response = Y
- Objective: Treatment effect of Zidovudine on mean CD4 counts



L(1)=high CD4 count, L(0)=Low CD4 counts

Stabilized weights calculation for k time points

$$SW_i = \frac{\prod_{k=0}^{K} P[A_k = a_{ki} \mid \vec{A}_{k-1} = \vec{a}_{(k-1)i}]}{\prod_{k=0}^{K} P[A_k = a_{ki} \mid \vec{A}_{k-1} = \vec{a}_{(k-1)i}, \vec{L}_k = \vec{l}_{ki}]}$$

It is the ratio of the probability (subject received the observed treatment given the past treatment) to probability (subject received his observed treatment given the past treatment and covariate history).

Ref: Hernan et. al, 2002

$(A(0), L(1), A(1))$	N_O	$E[Y\|\bar{A},\bar{L}]$	W	N_W	$f[A(1)\|A(0)]f[A(0)]$	SW	N_{SW}
(1,1,1)	15	100	4	15	0.39	1.56	23.4
(1,1,0)	15	100	4	15	0.11	0.44	6.6
(1,0,1)	63	90	2.22	35	0.39	0.87	54.6
(1,0,0)	7	90	20	35	0.11	2.20	15.4
(0,1,1)	5	100	4	5	0.43	1.72	8.6
(0,1,0)	5	100	4	5	0.07	0.28	1.4
(0,0,1)	81	90	2.22	45	0.43	0.95	77.4
(0,0,0)	9	90	20	45	0.07	1.40	12.6

(.63+.15)*.5

Where S(W)= $\dfrac{f[A(1)\,|\,A(0)]f(A(0))}{f[A(1)\,|\,L(1),A(0)]f(A(0))}$

$$\Pr[A(1)=1|L(1)=1,A(0)=1]=0.5 \text{ and } \Pr[A(0)=1]=0.5$$

$$W=1/(0.5\times 0.5)=4$$

Causal effect estimation

Causal effect = $E(Y_a - Y_{a'})$

Where a =(a(0),a(1))=(1,1) and

a`=(a(0),a(1))=(0,0)

First and third row are the subjects treated for the duration of the trial and 6 & 8th are untreated. Standard approach, weights the average by the observed number.

$E(Y_a)$= (15 x 100 + 63 x 90) / (15+63)

= 91.92

$E(Y_{a`})$= (5x100 + 9 x 90) / (5+9) =93.57

Therefore, causal effect=91.2-93.57= -1.65

This is biased because L(1) is a confounder on the effect of A(1) on Y

MSM approach using stabilized weights

$$E(Y_a) = \frac{23.4x100 + 54.6x90}{23.4 + 54.6} = 93.0$$

$$\text{Similarly, } E(Y_{a'}) = \frac{1.4x100 + 12.6x90}{1.4 + 12.6} = 91.0$$

Causal Effect= 2.0

- The IPTW estimate of 2.0 is the estimate of $\beta_1 + \beta_2 + \beta_3$ in the saturated MSM
  $E[Y_a] = \beta_0 + \beta_{1a0} + \beta_{2a1} + \beta_{3a0a1}$

- It can be shown that the non-stabilized and stabilized estimates will be the same in any saturated model.

(Refer to Hernan et. al, 2002)

Double Robust (DR) analysis

- The Double Robust estimator is based on the estimating function approach to MSM estimation.
- DR estimation is based on a model for both nuisance parameters: $Q(A,W) = E(Y \mid A,W)$ and $g(A \mid W)$

DR estimator of the MSM parameter

In practice, the DR estimator of the MSM parameter can be implemented in a three-step procedure:

- Estimate $Q(A,W) = E(Y \mid A,W)$
- Estimate $g(A \mid W)$
- Solve the DR estimating equations with the Newton- Raphson algorithm using the estimates from Q and g.

DR Estimation

DR estimation relies on two assumptions:

- The randomization assumption (or no unmeasured confounders): $A \perp Y_a \mid W, \forall a$
- Consistent estimation of one only of the nuisance parameters: $Q(A,W)$ or $g(A \mid W)$
- If g and Q are consistently estimated then the DR estimator is locally efficient, i.e. in particular more efficient than the IPTW estimator and as efficient as the G-computation estimator.

Targeted MLE (Van der laan and Rubin).

- The Targeted Maximum Likelihood Estimator (TMLE) is based on the maximum likelihood principle and can be linked to the DR estimating approach to MSM estimation to combine the attractive properties of both approaches
- TMLE estimation is based on a model for both nuisance parameters: $Q(A,W) = E(Y \mid A,W)$ and $g(A \mid W)$.

Targeted Maximum Likelihood

- MLE- aims to do good job of estimating whole density
- Targeted MLE- aims to do good job at parameter of interest
  - General decrease in bias for parameter of Interest
  - Fewer false positives

TMLE Implementation

- In practice, the TMLE estimator of the MSM parameter can be implemented in a three-step procedure:
- Estimate Q(A,W) = E(Y | A,W) (Qn(A,W))
- Estimate g(A | W) (g_n(A | W))
- An interative two-step procedure which can converge in one iteration:
  - Update Qn: regress Y on h(A,V) / g(A|W) with Qn as an offset.
  - Update βn: regress Qn on A and V according to the MSM, i.e.:

$$\beta_n = \text{argmin}_\beta \sum_i \sum_a (Q_n(a, W_i) - m(a, V_i \mid \beta))^2$$

TMLE contd…

- TMLE estimation relies on two assumptions:
  - the randomization assumption (or no unmeasured confounders):

    $A \perp Ya \mid W, \forall a$
  - Consistent estimation of one only of the nuisance parameters: Q(A,W) or g(A | W)

- Thus in a clinical trials where the model for g is guaranteed to be correctly specified, the inference is always valid and one can gain in efficiency.

- The TMLE estimator is easy to implement and the likelihood can be used to chose the best estimates when multiple solutions of the DR estimating equations exist.

Summary

- Causal Inference methodologies are very useful in Clinical Trials to improve efficiency.
- Targeted approaches (IPTW or T-MLE estimation and marginal structural models) can provide effect estimates in settings where
  - Randomization is not available (example: Drug safety)
  - Standard approaches can give biased estimates in the presence of confounding and intermediate variables (longitudinal data)

Additional References

- Robins, J.M., Hernán, M. and, Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5):550-560.
- Hernán, M., Brumback, B., and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology, 11(5):561-570.
- Van der Laan, M.J., and Robins, J.M. 2003. *Unified methods for censored longitudinal data and causality*. New York : Springer-Verlag.
- M.J. van der Laan. Causal effect models for intention to treat and realistic

  individualized treatment rules. Technical report 203, Division of Biostatistics,

  University of California, Berkeley, 2006a.
- M.J. van der Laan. Statistical inference for variable importance. *International*

  Journal of Biostatistics, 2(1), 2006b.
- M.J. van der Laan. Identity for npmle in missing data and biased sampling

  models. *Bernoulli*, 1(4):335–341, 1995a.
- Mortimer, K. Neugebauer, R. van der Laan, M. and Tager, I. 2005. An Application of Model-Fitting Procedures for Marginal Structural Models. *Am. J. Epidemiology* **162**:382–388.

Thank you!

Contact Info:

Thamban.Valappil@fda.hhs.gov

(301) 796-0828