Causal inference using potential outcomes: introduction and new challenges

Constantine E. Frangakis, PhD

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
cfrangak@jhsph.edu

Thanks to:
Don Rubin, Fabrizia Mealli

Outline

1. Framework of potential outcomes
2. Estimating effects of directly controlled factors
   (a) Role of models for potential outcomes
3. Estimating effects of directly controlled factors
   (b) Role of propensity score
4. Studies with noncompliance - instrumental variables
5. Partially controlled studies

Chapter 1. Framework

Section 1. What questions we regard as causal ?

Intervention.

Phrase 1. John can walk again because he had his hip replaced.

Phrase 2. Lucia survived breast cancer until her 90s because she got screening early.

Phrase 3. Mary is conservative because she is old.

Causal question: one that we can formulate with an explicit intervention

Section 2. Potential outcomes
Neyman (1923), Rubin (1974, 1977, 1978).

Subject: the speaker

Treatment $z = \begin{cases} 1 & \text{milk} \\ 0 & \text{no milk} \end{cases}$

Outcome $Y = \begin{cases} 1 & \text{asleep at 5:00 am} \\ 0 & \text{awake} \end{cases}$

Potential Outcomes: $Y(\text{milk})$ $Y(\text{no milk})$

Causal effect: comparison

Note:
Definition is separate from probability models.

To learn more we need more units.

	$Y(0)$	$Y(1)$	causal effect $Y(1)-Y(0)$
Subject 1			
2			
.	.	.	.
.	.	.	.
N			
of interest:	$\overline{Y}_0$	$\overline{Y}_1$	$\overline{Y}_1 - \overline{Y}_0$ $\overline{Y}_1/\overline{Y}_0$

$(Y_i(0), Y_i(1)), \quad i = 1, ..., N$

Each subject gets assigned one treatment.

Z	X	$Y(0)$	$Y(1)$
1		?	
0			?
0			?
1		?	
1		?	
0			?
		$\overline{Y}_0$	$\overline{Y}_1$

Z_i: treatment for subject i (assignment)

$\underline{Y} = (Y_i(0), Y_i(1))$

$Y_i^{obs} = Y(Z_i)$

X_i: covariates

pr$(\underline{Z}|\underline{Y}, \underline{X})$: assignment mechanism

Causal Inference as Missing Data problem

Z	X	$Y(0)$	$Y(1)$
1		?	
0			?
0			?
1		?	

Rubin-Causal-Model (Holland, 1986)	Before
$(Y_i(0), Y_i(1))$	$Y_i^{obs} = Y(Z_i)$
X_i: covariates	
Z_i: assignment for subject i	
pr$(\underline{Z}\|\underline{Y}, \underline{X})$: assignment mechanism	

Example: Randomizing subjects can be represented with potential outcomes but not with observed outcome.

$Z_i \coprod (Y_i(0), Y_i(1))$ (correct) $\qquad\qquad$ $Z_i \coprod Y_i^{obs}$ (incorrect)

Note: $\coprod$ means “independent of”

Section 3. Modes of inference

(i) **Fisher :** Statistic, its p-value for null.

(ii) **Neyman :** Statistic, its frequency properties under non-null.

(iii) **Likelihood and Bayesian :** Methods for creating procedures.

Common Goal: Inference on potential outcomes $(Y_i(0), Y_i(1)), i = 1, ..., N$.

What likelihood inference needs
(Rubin, 1978)

(i) Assignment pr$(Z \mid Y(1), Y(0), X)$ necessary.

(ii) Models for pr$(Y(1), Y(0), X \mid \theta)$:

- can think of probability as induced by sampling from a large population of units $(Y(1), Y(0), X)$

- Parameters θ contain causal effects

Next, we distinguish between studies where:

- The treatment of interest is directly controlled (intervened).

- The treatment of interest is different from the controlled factors.

Chapter 2. Effects of directly controlled treatments
(a) Role of outcome models

Section 1. Types of studies and assumptions

We consider studies in which:

1. we are interested in causal effects in the larger population of patients that the study sample represents
2. we have measured all variables that were used for assignment of treatments.
   (e.g., we can talk to doctors to find out what variables they measure before they recommend treatments)

Formalizing the above assumptions:

A.1. The study subjects $i = 1, ..., n$ are a simple random sample from a reference population, namely $\{Y_i(1), Y_i(0), X_i\}$ are "iid" from θ.

A.2 The assignment mechanism to treatment $(Z = 1)$ or control $(Z = 0)$ satisfies:

$$\text{pr}(Z_i = 1 | Y_i(1), Y_i(0), X_i = x) = e(x)$$

Note: this assignment is also called strongly ignorable
(Rosenbaum and Rubin, 1983).

Strong ignorability can be interpreted as if "there is randomization within levels of X".

Section 2. Estimation of causal effects under strong ignorability.

Goal: a causal effect, i.e., a comparison between $Y_i(0)$ and $Y_i(1)$ in a common set of subjects. So:

A causal effect is a function $Q(\theta)$ of the parameters θ of the distribution pr$(Y(0), Y(1), X \mid \theta)$. So:

Estimation of Q follows from estimation of θ

Assume the model

$$\text{pr}(Y_i(z) = y | X_i = x, \theta) = f(x, z, y, \theta)$$

Then, the likelihood of the observed data $\{Y_i^{obs}, Z_i \mid X_i\}$ is

$$\begin{aligned}
&\prod_i \text{pr}(Y_i^{obs}, Z_i \mid X_i, \theta) = \prod_i \text{pr}(Y_i^{obs} | X_i, Z_i, \theta) * \text{pr}(Z_i | X_i, \theta) \\
&= \prod_{i:Z_i=1} \text{pr}(Y_i^{obs} | X_i, Z_i = 1, \theta) \prod_{i:Z_i=0} \text{pr}(Y_i^{obs} | X_i, Z_i = 0, \theta) \prod_i \text{pr}(Z_i | X_i, \theta) \\
&= \prod_{i:Z_i=1} \text{pr}(Y_i(1) = y | X_i = x, \theta)_{y=Y_i^{obs}} \prod_{i:Z_i=0} \text{pr}(Y_i(0) = y | X_i = x, \theta)_{y=Y_i^{obs}} \prod_i \text{pr}(Z_i | X_i, \theta) \\
&= \prod_{i:Z_i=1} f(X_i, z = 1, Y_i^{obs}, \theta) \prod_{i:Z_i=0} f(X_i, z = 0, Y_i^{obs}, X_i, \theta) \prod_i e(X_i)^{Z_i} (1 - e(X_i))^{(1-Z_i)}
\end{aligned}$$

Note 1. The model of the potential outcomes, $\text{pr}(Y_i(z) = y | X_i = x, \theta)$, becomes the same as the model for the data, $\text{pr}(Y_i^{obs} = y \mid Z_i = z, X_i, \theta)$.

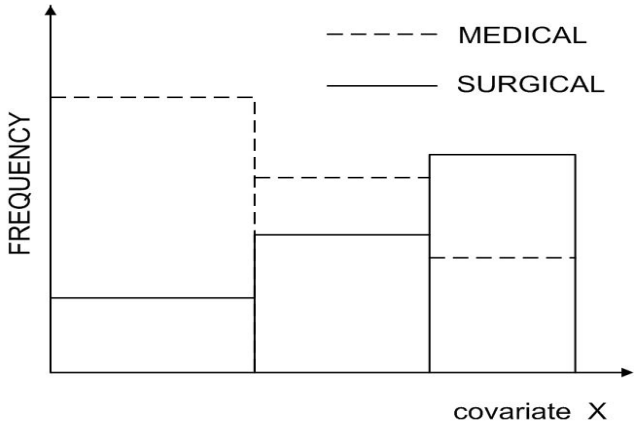
Note 2. Under “ignorability”, the probabilities of assignment $e()$ are ignorable.

Case study on patients with coronary artery disease

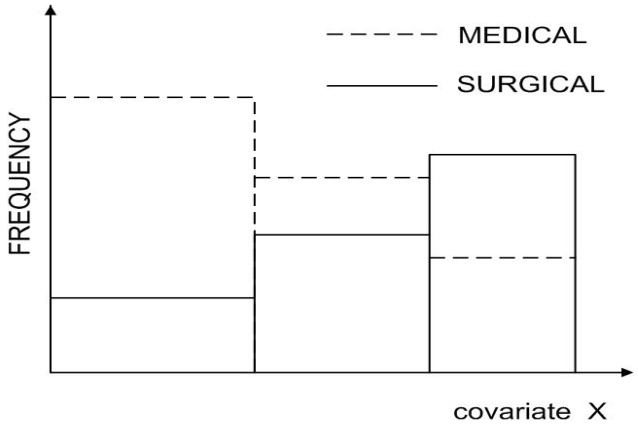
(Rosenbaum and Rubin, JASA, 1984)

- compare surgical: $z = 1$ vs. medical: $z = 0$ treatment; on functional ability Y at 6 months.
- Treatment recommendations mostly made by doctors.
- Number of variables ≈ 74.

Section 3. The case of single discrete X.



- goal here: $Q = E(Y_i(1) - Y_i(0))$
- discrete X covariate (e.g., risk index)
- we are assuming $(Y_i(1), Y_i(0)) \coprod Z_i \mid X_i$



- Fact:

  $$E\{Y_i(1)\} = \sum_k E(Y_i(1) \mid X_i = k)\, E(X_i = k)$$

- Data:

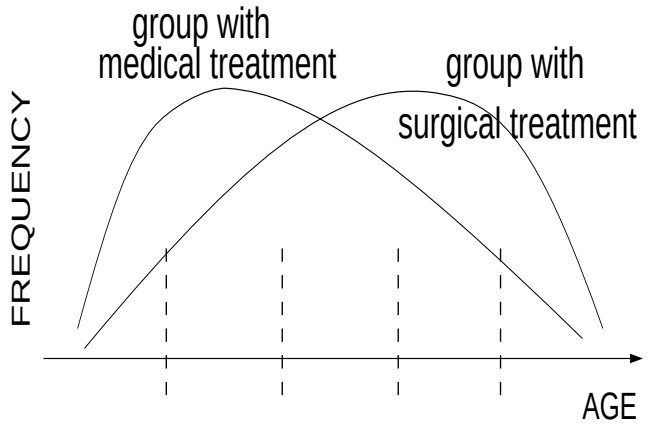
  n_k number of people in class $X_i = k$

  $\bar{y}_{z,k}$ observed average Y in class k
  among those assigned treatment z

From likelihood, under no further assumptions:

MLE of $E\{Y_i(1) - Y_i(0)\} = \sum_k \{\bar{y}_{1,k} - \bar{y}_{0,k}\} \frac{n_k}{n}$

If X is continuous,



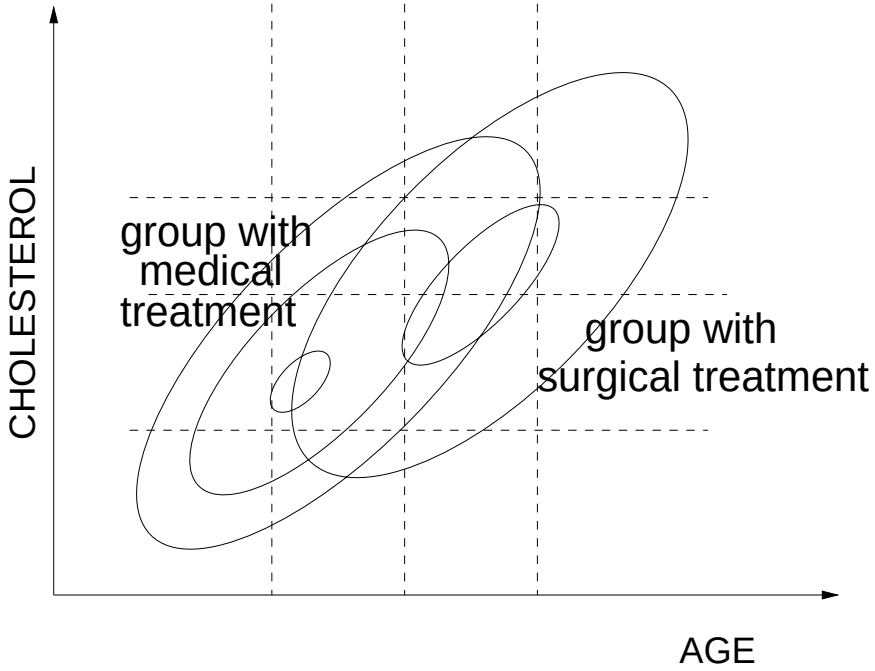
- can create K classes
- can get estimator $\hat{Q}^{(K)} = \sum_{k=1}^{K} \{\bar{y}_{1,k} - \bar{y}_{1,k}\} \frac{n_k}{n}$

Then for $K = 5 - 7$ classes, we very often have (Cochran, 1968)

$$\text{Bias Reduction} \quad 1 - \frac{\widehat{Q}^{(K)} - Q}{\widehat{Q}^{(1)} - Q} \approx 90\%$$

Notes

- with many covariates, direct sub-classification becomes infeasible.



- Model directly outcome on many covariates decreases reliability
- Search for a scalar summary of covariates that balances all of them.

Chapter 3. Effects of directly controlled treatments
(b) Role of the propensity score

Section 1. Definition and main properties

(def.) Propensity score is the probability the person gets one treatment (vs. another) given covariates. (Rosenbaum and Rubin, 1983).

$$e(x) = pr(Z_i = 1 \mid X_i = x)$$

Two key properties:

(i) Covariate balance property: Treatment and control subgroups with the same scalar e have the same distribution of all covariates entered in e. Namely

$$Z_i \coprod X_i \mid e(X_i)$$

(ii) If the assignment is ignorable given X, then it is also ignorable given only $e(X)$. Namely

$$\text{If} \quad (Y_i(1), Y_i(0)) \coprod Z_i \mid X_i \text{ then } (Y_i(1), Y_i(0)) \coprod Z_i \mid e(X_i)$$

Notes

- By ignorability property: a study that was "as randomized given multivariate X", can now be treated "as randomized given scalar $e(X)$".

- The balancing property allows us to check models for prop. score.

- Estimating $e(X)$ is a desirable trade-off to $(Y(z) \mid X)$:
  - if e(X) involves human rules on assignment that we can access (e.g., talk to doctors), and
  - if $(Y(z) \mid X)$ involves Natural laws we cannot easily access

Section 2. Estimation and use of propensity score.

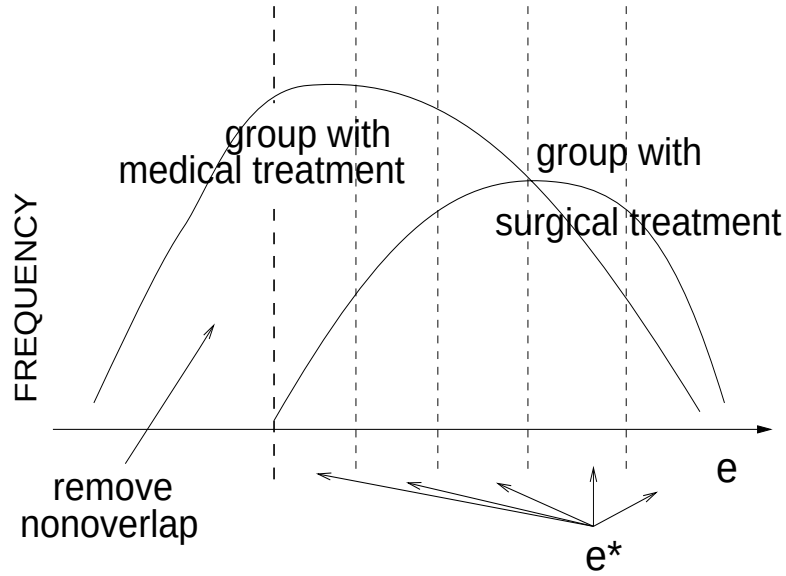
Talk to those who made asignment, to elicit X and broad rules. Then,

(i) Estimate $e(x) = pr(Z_i = 1 \mid X_i = x)$, for example with preliminary logistic model

$$\text{logit} \quad e(\underline{X}) = \text{ first order terms of } X \tag{1}$$

(ii) Plot histograms of e_i and remove nonoverlaps.

(iii) Subclassify the people based on 5-7 subclasses: e_i^*.



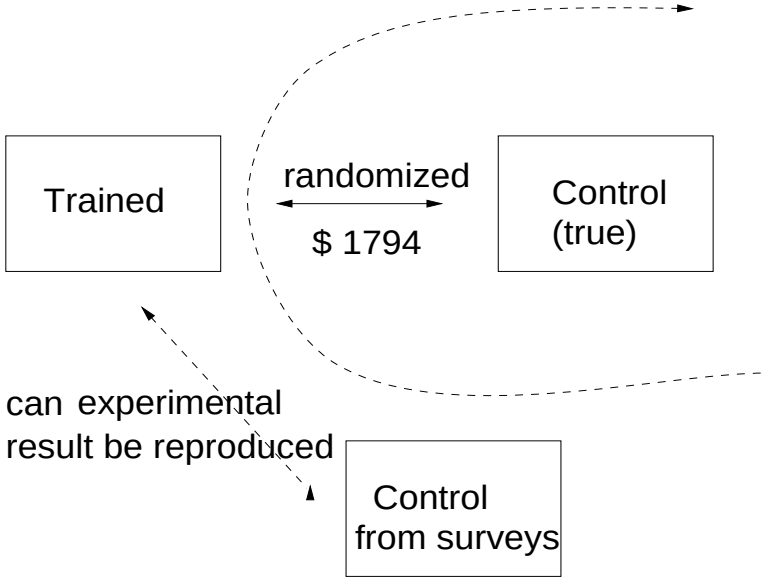
(iv) Check if classes balance covariates;
X_i independent of $Z_i \mid e_i^*$.

(v) If not, increase structure in (1) (e.g. higher order terms) of X's; repeat (i)-(iv) until satisfied by balance.

Then, use propensity classes for outcome comparisons
using likelihood principles we saw earlier (includes matching).

Section 3. Case study on job traininig
National Support Work Demonstration (1975).
(Lalonde, *Am. Econ. Review*, 1986), here subset of data

- 3000 unemployed randomized to (i) job training, (ii) control.
- Outcome Y = Earnings in 1978.
- Effect here = $ 1794



- "Hide" true control, and try to reproduce results from observational data.

– Using outcome regressions without propensity score: did not reproduce results.

– Using propensity score: better reproduced results.
(Dehejia and Wahba, JASA, 1999)

Chapter 4. Studies with noncompliance - instrumental variables

Section 1. Types of studies we consider

Studies where original assignment is controlled, but some participants take different treatment(s) from those originally assigned

Why we consider these studies ?

- Noncompliance arises often in clinical trials, and serves as template in other settings
- Noncompliance introduces new goals

Section 2. Impact of noncompliance on goals

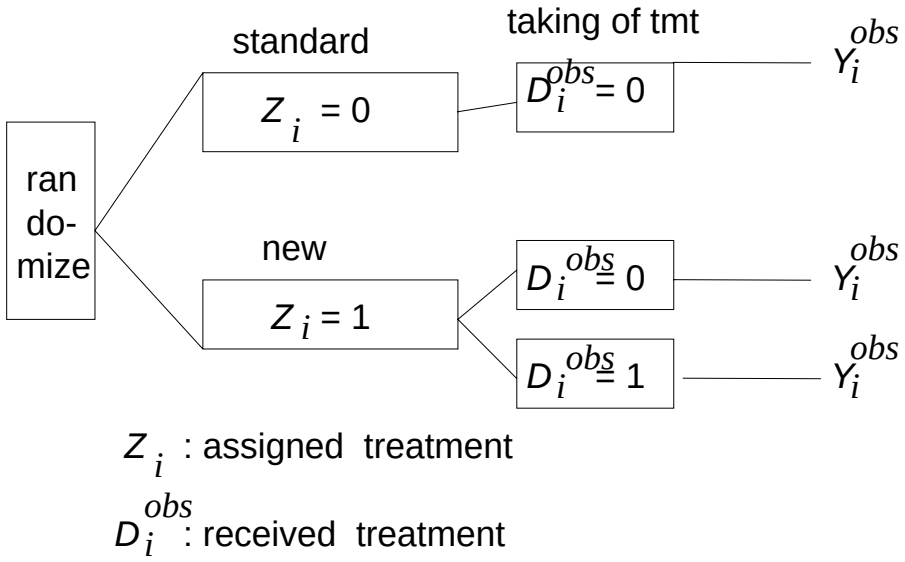
Case study: trial on Vitamin A (Sommer and Zeger 1991)

Assignment Z_i	Vitamin A taking D_i^{obs}	Mortality Y_i^{obs}	Number of Units (total 23,682)
0	0	0	11,514
0	0	1	74
1	0	0	2385
1	0	1	34
1	1	0	9663
1	1	1	12

Why care about effect of taking vitamin A ?

- If people are told the efficacy is high (low) they may change compliance.
- It is the biologic effect that will tell us if we need more research on role of Vitamin A in immune response.

Section 3. Standard definitions and problems

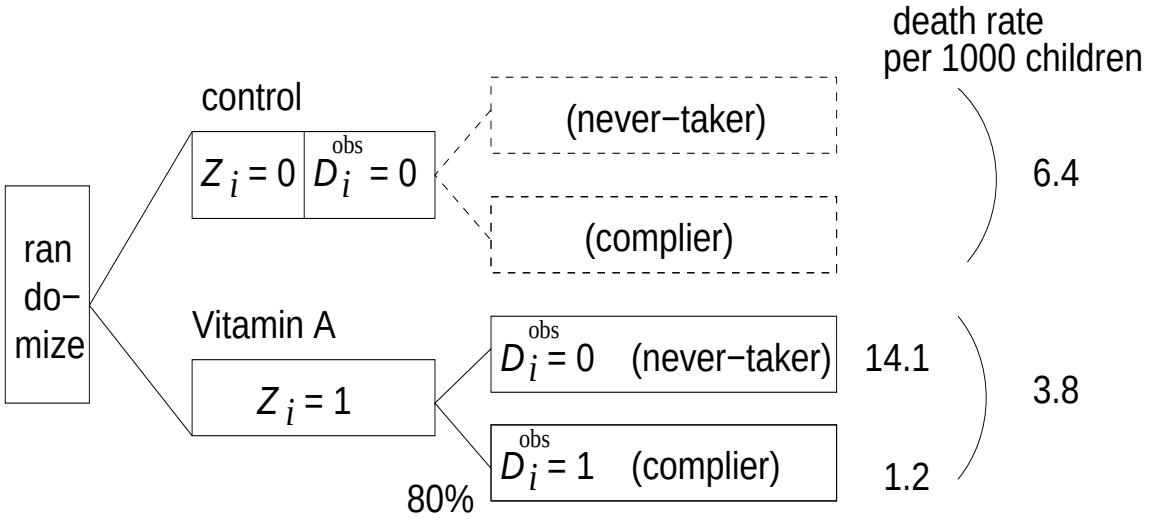


- Intention-to-trear (ITT) method.
  ($Z_i = 0$) versus ($Z_i = 1$).

- As-treated: ($D_i^{obs} = 0$) versus ($D_i^{obs} = 1$).
  Does not generally estimate a treatment effect.

- Per-protocol: ($Z_i^{obs} = 0$ and $D_i^{obs} = 0$) versus ($D_i^{obs} = 1$).
  like as-treated, not generally comparable groups.

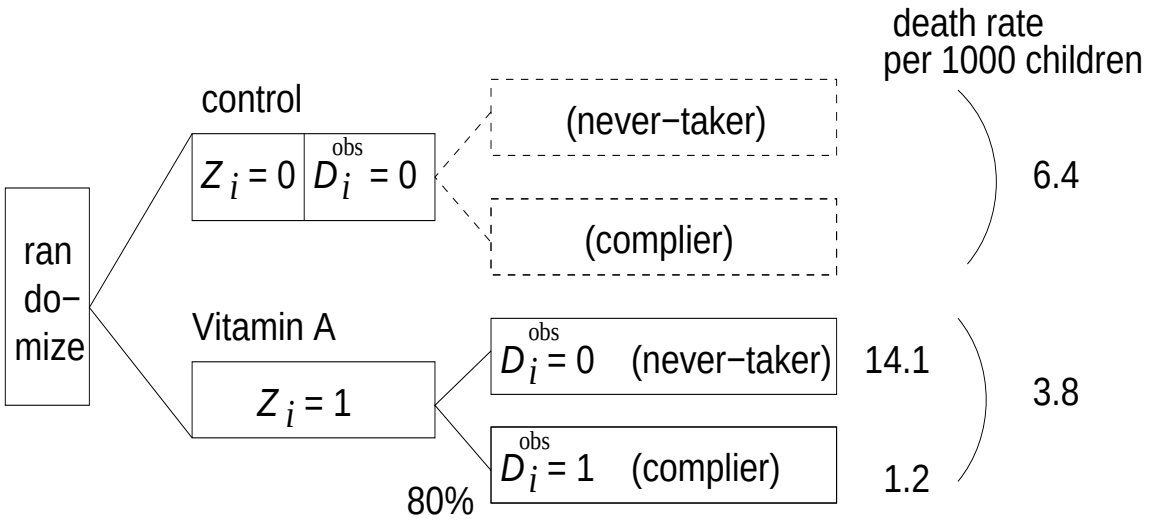
Section 4. The compliance status



Compliance status (Imbens and Rubin, 1997, Annals of Statistics)

Compliance Status (%)	potential receipt	
	$D_i(0)$	$D_i(1)$
"never-taker"(π_n)	0	0
"true complier"(π_c)	0	1

Section 5. Definition of efficacy using Compliance Status



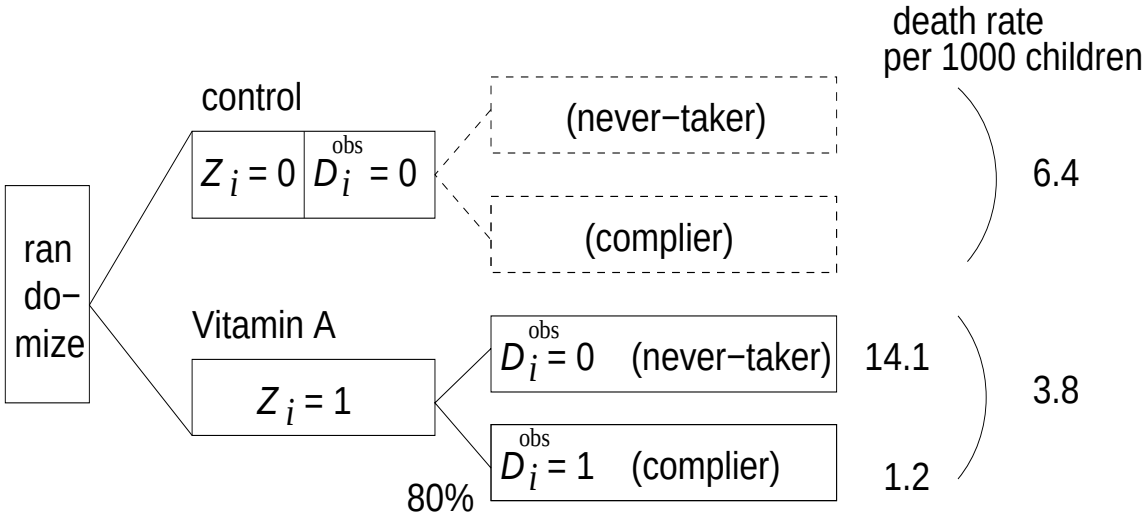
Note:

(a) Being a “true complier” does not change with assignment.

(b) The “true compliers” is the only group for which:

the comparison of outcomes under receving vs. not receiving treatment = the experimental comparison (of randomized arms).

Section 6. Estimation of efficacy using assumptions of instrumental variables



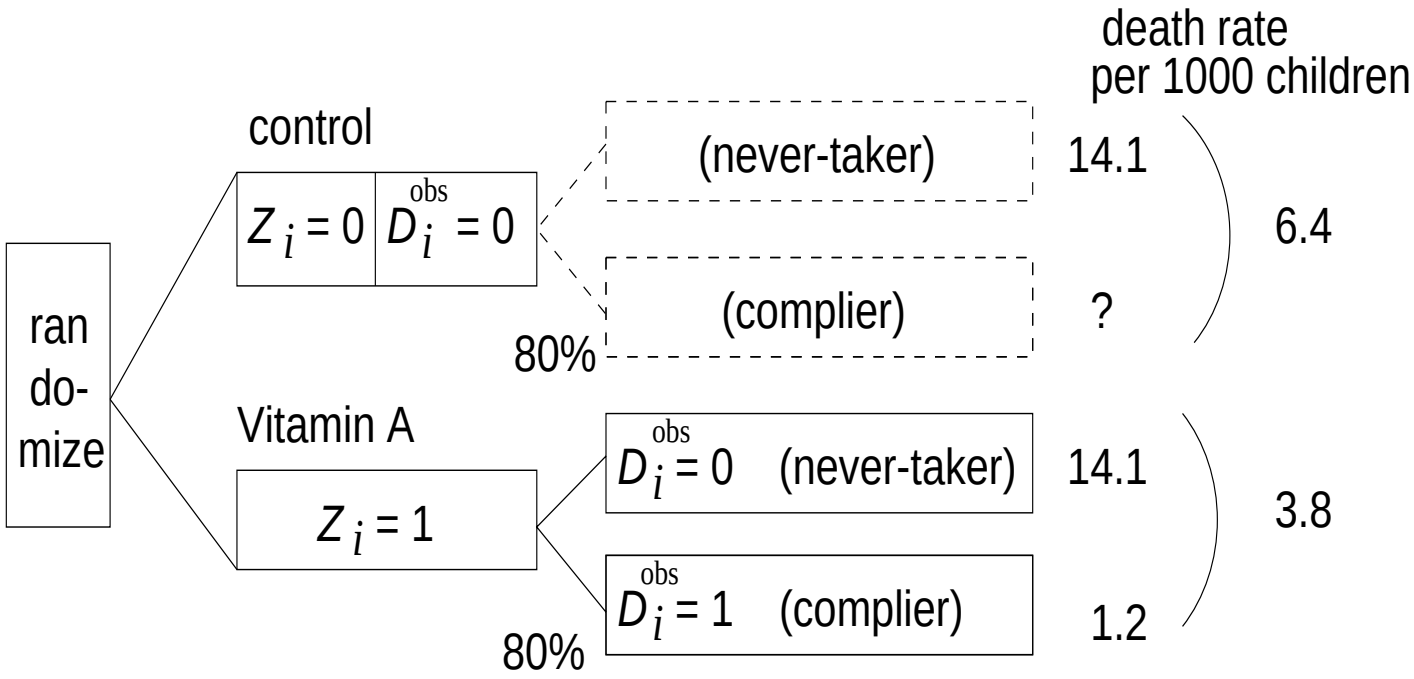
Assumptions:

– Ignorable assignment: $Z_i \coprod (Y_i(0), Y_i(1), D_i(0), D_i(1))$

– Exclusion restriction: If $D_i(0) = D_i(1)$ then $Y_i(0) = Y_i(1)$

Under above assumptions, the complier average causal effect is estimable

(Angrist, Imbens and Rubin, JASA, 1996)



$$6.4 = (14.1)\ 20\% + (?)\ 80\% \Rightarrow ? = 4.5$$

	all children	true compliers only
death rate if assigned control	6.4	4.5
death rate if assigned Vit. A	3.8	1.2
Decrease in death rate	2.6	3.3
Relative risk	0.59	0.27

Notes

- Spirit of assumptions is as in instrumental variables
- Framework of potential outomes clarifies implications of the different assumptions

Chapter 5. More general partially controlled studies

Section 1. Types of studies we consider

We consider studies with

- controlled factors z
- factors $D(z)$ of indirect control, by z
- outcome of interest $Y(z)$.

In such studies, we are often interested in comparing outcomes:

$$\{Y_i(z=1)\} \text{ to } \{Y_i(z=0)\} \text{ conditionally on strata of } (D_i(0), D_i(1))$$

We call strata $(D_i(0), D_i(1))$ "principal strata"

(Frangakis and Rubin, Biometrics, 2002)

Why we consider these studies ?

- They provide template to estimate effects of multiple factors that are not directly controlled
- Such factors introduce new goals (as with IV), but also new analyses and designs

Section 2. Multiple sources of protocol deviations – Impact on analyses

Often, there are protocol deviations of different types :

– non-compliance with treatment

– not providing outcome

In these cases even the estimation of the intention-to-treat effect is problematic

Frangakis and Rubin, *Biometrika*, 1999

Mealli and Rubin, *Journal of Econometrics*, 2003

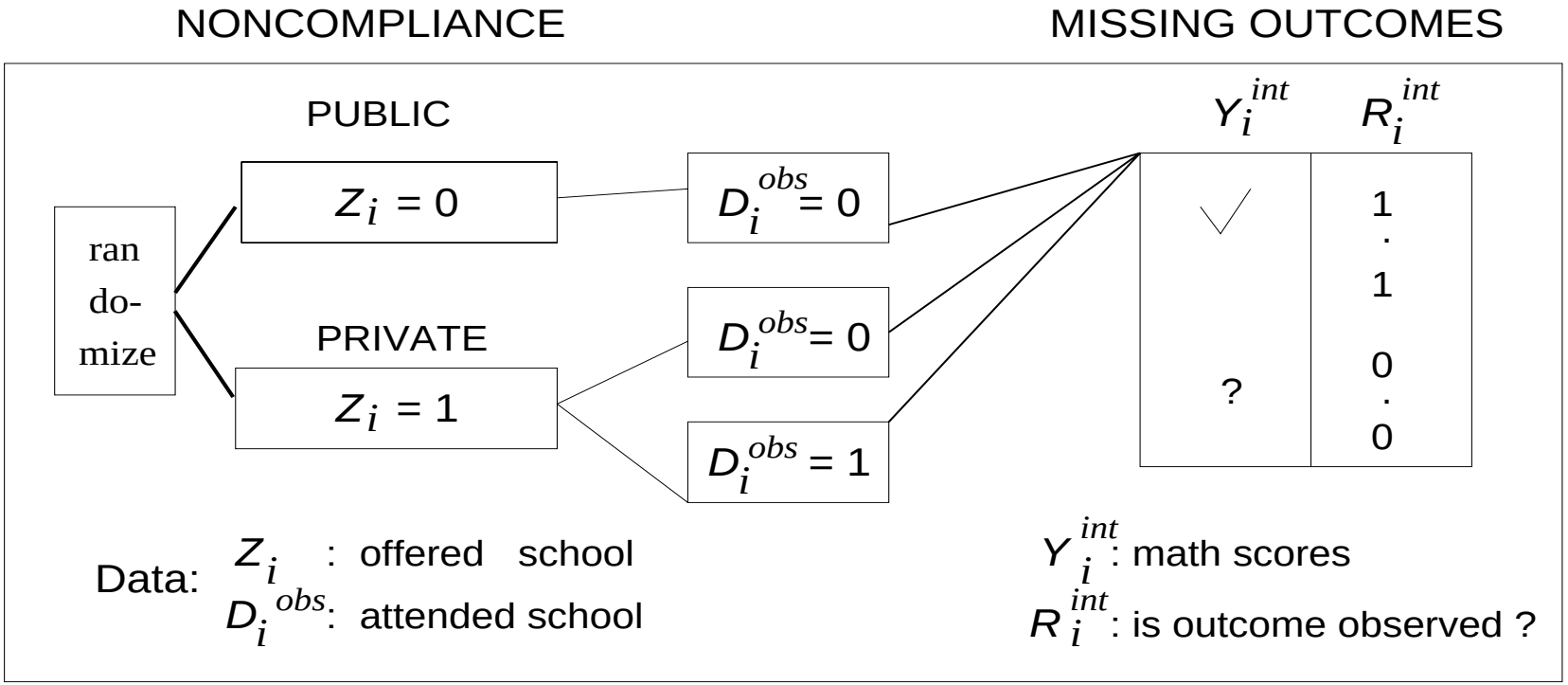
Frangakis et al., *JASA*, 2004

Mealli, *Biostatistics*, 2004

Case study: School choice randomized trial

(Barnard, Frangakis, Hill & Rubin, JASA 2003)

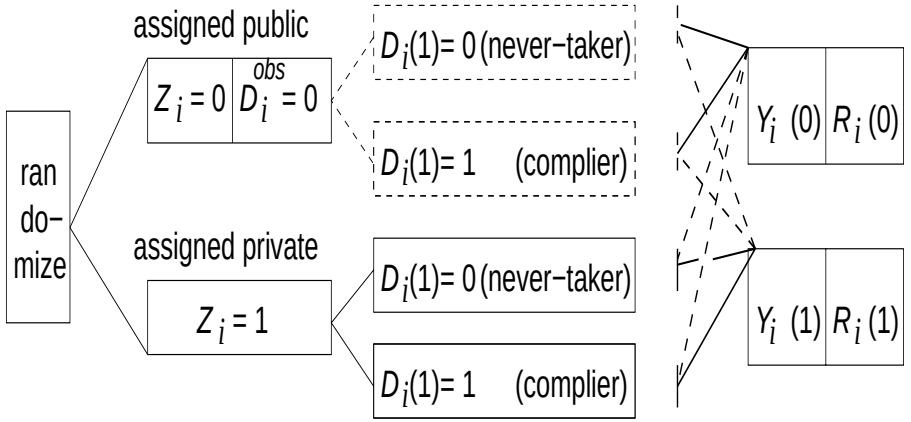
Goal: Comparing private ($Z = 1$) to public schools ($Z = 0$): math scores (also, drug abuse, incomes etc.)



Goal 1: Intention-to-treat (ITT) effect $E(Y_i(1)) - E(Y_i(0))$.

Goal 2: Effect on"compliers", $E(Y_i(1)|C_i = c) - E(Y_i(0)|C_i = c)$

Section 3. Standard ITT analyses to estimate ITT effect



Note on distinction:

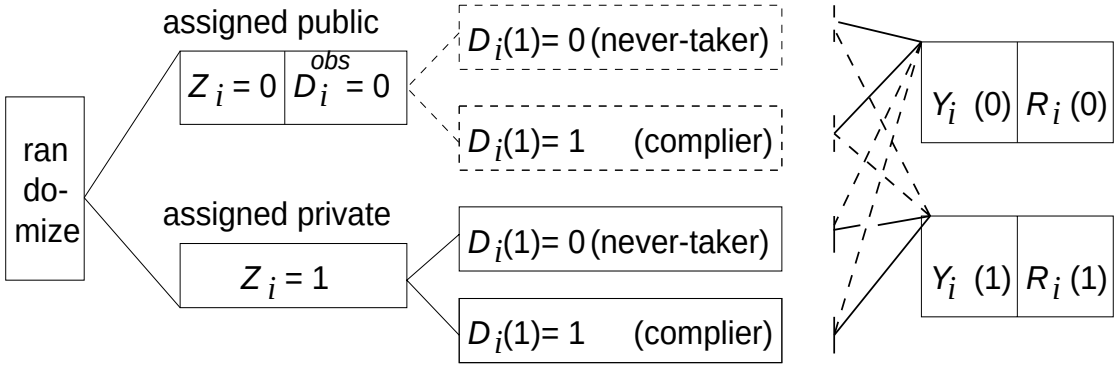
- Intention-to-treat effect: the effect that ignores compliance.

  $E(Y_i(1)) - E(Y_i(0))$

- Intention-to-treat method: a method that ignores data on compliance.

Validity of ITT method requires that:
available outomes be comparable to unavailable outcomes
(adjusting for obs variables)

Section 4. Principal strata of compliance and loss-to-follow-up



When the same person i :

is assigned :	the school attended is :	the (math) outcome is:	the loss to follow-up status is :
public	public	Y_i (0)	R_i (0)
private	D_i (1)	Y_i (1)	R_i (1)

With respect to school " compliance":

- children who would never attend private school in the study no matter the lottery assignment, $\{i : D_i(1) = D_i(0) = 0\}$, labeled "never-takers" and denoted by $C_i = n$; and
- children who would comply with either lottery assignment, $\{i : D_i(1) = 1$ and $D_i(0) = 0\}$, labeled "compliers" and denoted by $C_i = c$.

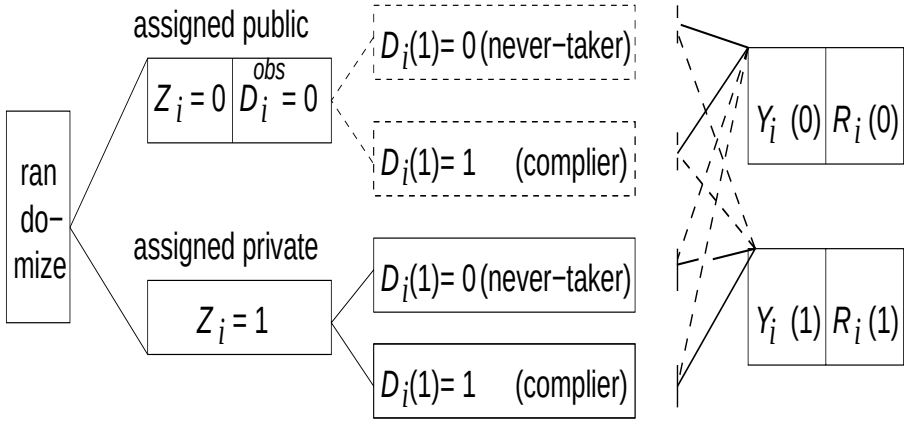
With respect to school " follow-up": similarly

"...an analysis based on groups defined according to whether or not they took the prescribed medication destroys ... balance ."

Farwell et al., *N. Engl. J. Med.*, 1990; 322: 364–9.

Compliance (latent) principal strata C_i are important.

Section 5. Assumptions allowing importance of principal strata



Assumption 1. Latent ignorability

Allows that principal strata have different outcomes, and loss to follow-up behavior.

$$Y_i(z) \text{ independent of } R_i(z) \text{ given } C_i$$

Assumption 2. Compound exclusion:

If person i is a never-taker, then

$$\begin{pmatrix} Y_i(z=1) \\ R_i(z=1) \end{pmatrix} = \begin{pmatrix} Y_i(z=0) \\ R_i(z=0) \end{pmatrix}.$$

Section 6. Invalidity of ITT analyses for estimating ITT effect

Under the above assumptions:

1. ITT standard methods (i.e. that ignore compliance data in the analysis) are inconsistent for ITT effect.
2. ITT effect is estimable (with no other assumptions) by an estimator that uses *compliance* and *follow-up behavior* data.

Frangakis and Rubin, *Biometrika* 1999

Mealli, *Biostatistics*, 2004

Section 7. Results from school choice

- Offer of vouchers (ITT effect) was estimated to increase scores only for math, for approx 3% ranks overall, with results relatively consistant across grades and type of school (new method).
- Estimates of more standard method were very variable across grades and type of school.

Section 8. Impact on designs

To evaluate a treatment/programme, increasingly studies cannot control the treatment, e.g., due to

- expense, or
- many “competing” treatments

In many such cases,

- we can design our evaluation as a partially controlled study
- we can define estimands and analyse the data using the framework of “principal stratification”

Remarks

The framework of potential outcomes

- provides a fundamental perspective for causal inference
- gives insights to better methods and designs for challenging evaluations

For references

Email: cfrangak@jhsph.edu, or

Go to: http://biosun01.biostat.jhsph.edu/∼cfrangak