Potential outcomes model:

Approach to defining effects of a cause/intervention

Define effect of treatment/exposure (e.g., effect of regular aspirin use)

Comparison of outcomes in a given individual/population under different levels of treatment/exposure

Examples:

Effect of Aspirin

Compare 1 year mortality that would occur in given population if

1) All took 1 aspirin/day

2) None took aspirin (except as analgesic)

Think of 2 hypothetical experiments

Compare survival curves that would be observed in subjects with acute appendicitis if

1) Provide appendectomy

2) Fail to provide appendectomy

Compare incomes after 5 years for subjects if

1) Draft into military service

2) Don't draft

Compare survival curves for subjects with sickle cell disease if

1) Gene producing sickling replaced

2) Gene not replaced; current standard of treatment

Sometimes, means of intervention unavailable/unknown

Can perform “thought experiment”

Questions:

About effect of intervention

whose answers are of interest to people making decisions regarding whether to intervene

Not about comparing treated and untreated subjects

Not directly observable in given population (without further assumptions)

If provide/take aspirin to subject, can’t observe directly what would have happened if didn’t provide

If don’t provide, can’t observe what would have happen if did provide

Notation:

$A \equiv$ treatment/exposure/intervention of interest

Use terms interchangeably throughout course:

Some distinctions between terms:

treatment: potentially beneficial, used in response to particular condition; term used in statistics/econometrics literature; reasons for use/application related to outcome under study (people get appendectomy to prevent fatal consequences of appendicitis)

exposure: reasons for use/exposure often not related to outcome under study; often no *a priori* reason to think beneficial

for course, distinctions somewhat arbitrary

Concentrate on binary *A*

A=1: treated/exposed
A=0: untreated/unexposed

Use subscript *i*: A_i is level of treatment received by individual *i*

Often dropped/implicit in notation

Outcome Y_i:

Can be

binary
 vital status at fixed point in time: Y=0 (1)- alive (dead)

continuous
 height, weight

failure-time/time to event (will sometimes use T)

multidimensional
 repeated measures (repeated binary)
 longitudinal data

will concentrate on 1-dimensional

Potential outcomes $Y_i^a \equiv$ outcome that would be observed in subject *i* if that subject were given treatment a

Y_i^1	vital status time to death height, weight	that would be observed if treated/exposed
Y_i^0	vital status time to death height, weight	that would be observed if untreated/unexposed

Implicitly, potential outcomes are well-defined even under treatments not received

For untreated subject, unobserved Y_i^1 is well-defined

For treated subject, unobserved Y_i^0 is well-defined

For treated subject, Y_i^0 is outcome that would have been observed if, contrary to fact, the subject had not been treated; Y_i^0 is thus a *counterfactual* variable or outcome

Observed outcome Y is potential outcome under treatment subject receives:

i.e., $Y = Y^A$

Link between potential outcomes and observables

Treatment determines which potential outcome observed

Missing data problem

Implicit assumptions in notation:

1. Independence: the treatment given one person does not affect outcome in another

   Potential outcomes formulation: $Y_i^{a_i, a_j} = Y_i^{a_i, a_j'}$ (i.e., potential outcome for subject *i* is the same no matter what treatment subject *j* given)

   Sometimes grossly violated

   Infectious disease epidemiology; vaccinating one person in household may reduce probability of others in household becoming infected, so $Y_i^{a_i, a_j=1} \leq Y_i^{a_i, a_j=0}$ for some subjects

   Problem: simpler notation with just 1 index treatment inadequate

   Extensions of ideas to infectious disease setting considered by Halloran and coauthors

2. Potential outcome well-defined/stable/consistency assumption

In hypothetical repetitions of an experiment on any individual, outcome would be same

Consider hypothetical trial of effect of free alcohol on death due to automobile collisions:

If provide subject alcohol at 1 AM, might run into oncoming vehicle; at 1:01 AM, may be no oncoming vehicle, so will live

Outcome under exposure level: provide alcohol not well defined

Assumptions 1 and 2: Stable Unit Treatment Value Assumption (SUTVA; Rubin)

Consistency assumption

Can circumvent problems of being ill-defined by concentrating on distributions of potential outcomes (see Hernan and Robins); stochastic potential outcomes/counterfactuals

What are possible meanings of stochastic potential outcomes?

Stochastic world (as in quantum mechanics)

impossible to fully know future

Deterministic world

use probability to represent our uncertainty

e.g., inability to fully specify intervention

average over presumed uncertainty about this

What is relation to views of probability and statistics?

parallels two different views of probability

- long-run average frequencies
  - model underlies standard/frequentist statistics
- subjective statements about beliefs
  - model underlies subjective Bayesian approach to statistics/prior distributions

Causal effects: comparisons (contrasts) of potential outcomes for individual or group

For binary treatment

Comparison of Y^1 and Y^0

For binary outcome, 4 causal types (Greenland and Robins, 1986):

Causal type (T)	Y^0	Y^1	$Y^1 - Y^0$ ($E(T)$)
1. Treatment fatal	0	1	1
2. Always live	0	0	0
3. Always die	1	1	0
4. Treatment curative	1	0	-1

For other kinds of outcomes, tedious or impossible to provide exhaustive categorization of causal types

Can consider measures of effect other than differences:

Suppose Y is failure-time outcome:

Y^0: time to failure if not treated
Y^1: time to failure if treated

Measures of effect:

$Y^1 - Y^0$: Lengthening in time (in days, weeks, etc.) to failure

$\frac{Y^1}{Y^0} - 1$ Proportional lengthening in time to failure (no units)

Effect of treatment received:

Contrast of $Y = Y^A$ and Y^0

Consider subjects of type: treatment fatal ($Y^0 = 0$, $Y^1 = 1$)

A	Y	Effect of treatment received
0	0	Treatment would have killed
1	1	Treatment killed

Implicit comparison level (no treatment)

Actual vs. potential effect

Analogous to attributable fraction vs. attributable fraction in exposed

Effects in populations-2 types:

1. Comparison of distributions of potential outcomes in population (population averaged)

2. Distributions of contrasts in potential outcomes in population (subject-specific)

use of terms not same as in (say) categorical data analysis

Consider effect types for binary outcome:

1. Compare $pr(Y^0=1)$ with $pr(Y^1=1)$

   Risk differences: $pr(Y^1=1)-pr(Y^0=1)$

   Risk ratios: $pr(Y^1=1)/pr(Y^0=1)$

   Odds ratios: $odds(Y^1=1)/odds(Y^0=1)$

2. Proportion of subjects with causal types T=1,....,4 (or functions of these proportions)

   Proportion of types: $pr(T=t)$

   Weighted sum (average) of individual effects: $\sum_t pr(T=t)\delta(t)=E(Y^1-Y^0)$

For difference measures of effect, difference in average outcomes equals average difference in potential outcomes (average individual effect):

$$E(Y^1) - E(Y^0) = E(Y^1 - Y^0)$$

Not true for other measures of effect (e.g., risk ratios, odds ratios)

For other outcomes, can compare distributions or functions of distributions

Potential outcomes model: conceptual model for defining causal effects

Restrictions on data/estimation not considered

How does potential outcomes model relate to standard missing data problem?

What is unique about causal inference setting? (Hint: consider what is observable)

Cannot observe simultaneously both Y^0 and Y^1; never have complete data

How might this influence what can be estimated? (Consider this in homework problem)