



JOHNS HOPKINS  
BLOOMBERG SCHOOL  
*of* PUBLIC HEALTH

# Methods in Drug Effectiveness Research: Part 2

---

Jodi Segal, MD, MPH  
Johns Hopkins University



Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Outline

1. Review of some common biases
2. Advantages of new user and active comparator designs
3. Immortal time bias
4. Graphical depiction of study design
5. Review of confounding
6. Propensity score methods
7. Effectiveness trials





# Review of Some Common Biases

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Review of Common Biases in Pharmacoepidemiology

- ▶ Why discuss ...
- ▶ Good decisions at the time of **design** of a study can reduce bias

# Review of Bias

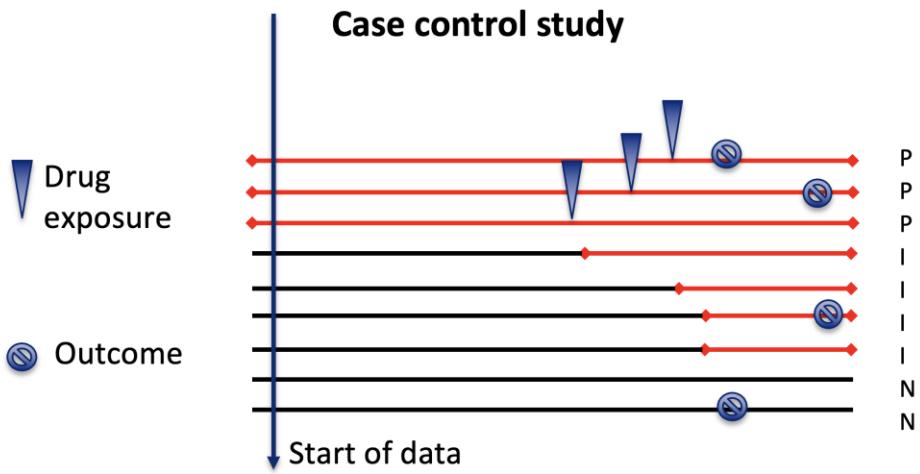
- ▶ Bias is a systematic error in the data
- ▶ Bias consistently pulls the risk estimate away from its true value



# Selection Bias: Prevalence Bias Is a Type of Selection Bias

- ▶  $\text{Prevalence} = \text{incidence} \times \text{duration}$
- ▶ Prevalent cases may have a longer duration of disease and more *opportunity* to have been exposed to the drug of interest
- ▶ It could even be that the drug is prescribed and prolongs life (and disease) and thus looks like drug is associated with *disease*
- ▶ This is most common in **case-control studies**

# Selection Bias: Prevalence Bias



P = prevalent cases of disease

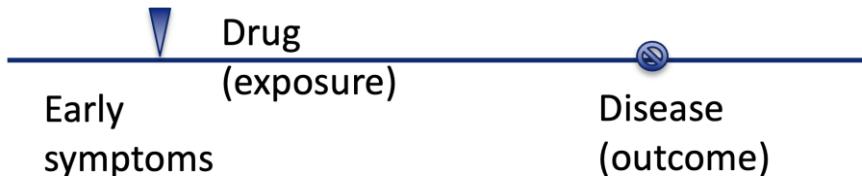
I = incident cases of disease

N = no disease

- If only incident cases (I) in the period of interest had been included, there would have been no association between drug and outcome (the incident cases weren't sick long enough to have exposures)

# Selection Bias: Protopathic Bias Is a Type of Selection Bias

- ▶ This is a term unique to pharmacoepidemiology
- ▶ **Protopathic bias** may occur “if a particular maneuver [drug] was started, stopped, or otherwise changed because of the baseline manifestation caused by a disease or other outcome event”
- ▶ Occurs when a drug is prescribed for an early manifestation of a disease that *has not yet been diagnostically detected*—think about early behavioral symptoms associated with dementia
- ▶ When the disease is later diagnosed, a causal relationship may be incorrectly inferred between the drug and the disease



Source: Horwitz, R. I., & Feinstein, A. R. (1980). The problem of “protopathic bias” in case-control studies. *The American Journal of Medicine*, 68(2), 255–258.  
Image source: Jodi Segal.

# Information Bias: Misclassification Is a Type of Information Bias

- ▶ Misclassification occurs because of errors in measurement or recording of the disease or exposure
- ▶ Nondifferential misclassification:
  - ▶ Misclassification between groups is approximately equal
- ▶ Differential misclassification:
  - ▶ Amount of misclassification *differs* between groups (“systematically biased”)

# Information Bias

*ICD = International Statistical Classification of Diseases*

## Information bias in pharmacoepidemiology studies

- ▶ Information bias is very common in pharmacoepidemiology studies
  - ▶ Claims data: uncertain sensitivity and specificity of codes (ICD-10) in a given data set
  - ▶ Medical records data: uncertain sensitivity of diagnoses in physician records (incomplete recording of diagnoses)
  - ▶ Survey data: incomplete recording of sensitive information
- ▶ Happily, the misclassification is *often* nondifferential; the misclassification occurs to the same extent across the exposure groups or across the outcome groups

# Differential Misclassification Certainly Can Happen

- ▶ Recall bias
  - ▶ Cases and controls may have different memories of their past drug exposures
- ▶ Detection bias
  - ▶ More attention to exposure assessment in cases compared to controls
  - ▶ Different follow-up procedures in cohort studies according to exposure

# Differential Misclassification: Detection Bias Example

- ▶ What would this look like in a cohort study?
- ▶ Study of individuals with acne:
  - ▶ Those on Accutane are asked to come every month for evaluation
  - ▶ Those on antibiotics are asked to come yearly for evaluation
  - ▶ More frequent contact will result in more *detection* of adverse events
- ▶ Study of individuals with hypertension in which claims data is used:
  - ▶ Those taking amlodipine have many office visits for leg swelling (a common side effect) and they proceed to echocardiography
  - ▶ Those taking losartan do not come in often and do not have echocardiography
  - ▶ More intensive evaluation will result in more diagnoses of mild heart failure in amlodipine group because of more *detection*

# Possible Solutions to Misclassification Biases

- ▶ Blinding of data collector to exposures (if collecting new data)
- ▶ Standardization of ascertainment processes
- ▶ Objective definitions for exposures and outcomes
- ▶ Very challenging when using existing (“real-world”) data



# Advantages of New User and Active Comparator Designs

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.

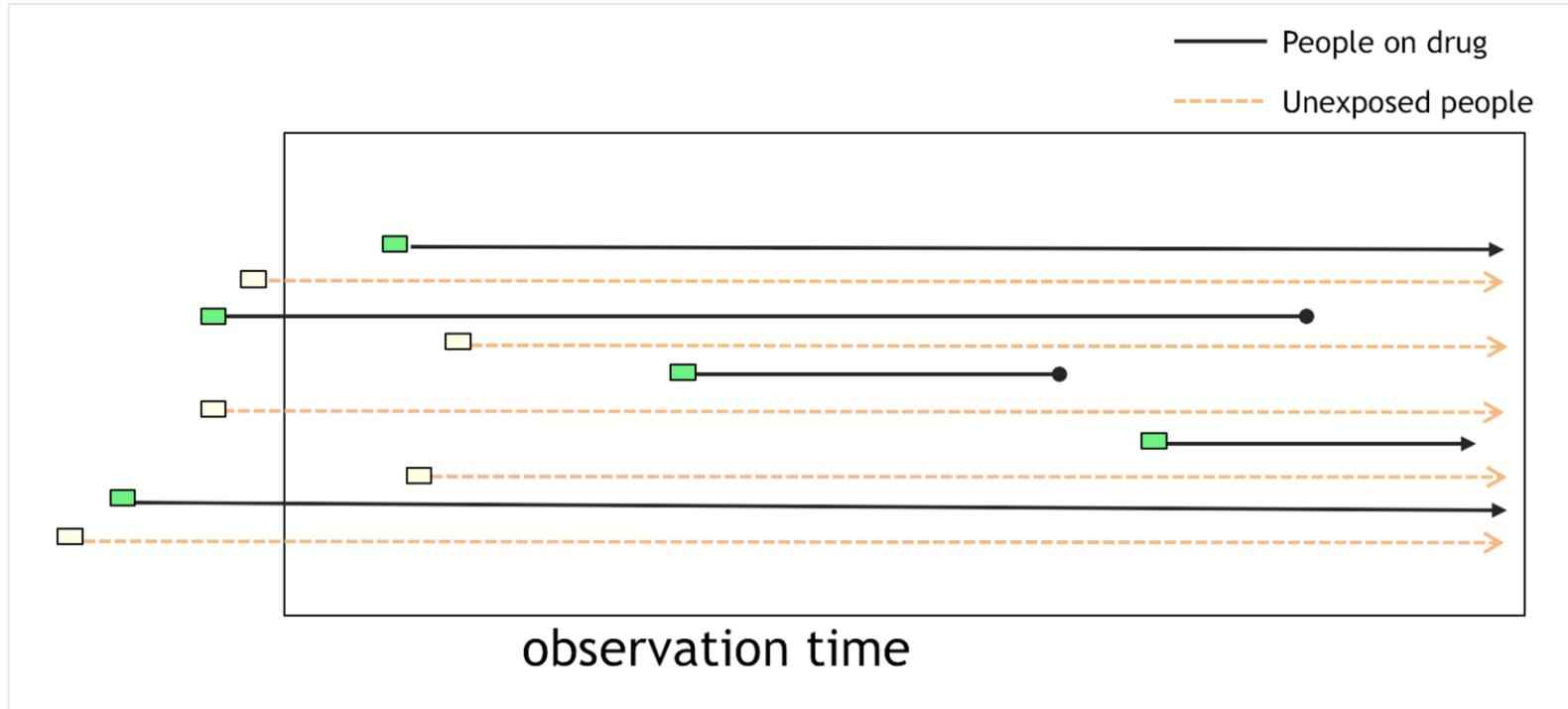
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Advantages of New User and Active Comparator Designs in Pharmacoepidemiology

- ▶ Good design helps us overcome some biases in cohort studies
- ▶ Makes us think about how we define the eligible population

| <b>Prevalent users of drug</b>             | <b>Incident users of drug (new users)</b> |
|--|---|
| People already taking the drug of interest | People newly taking the drug of interest  |

# Problematic Mix of Incident and Prevalent Drug Users



# Challenges With Including Prevalent Users of Drug

- ▶ Under-ascertainment of events that occur early after start of exposure
- ▶ Study population is people who have demonstrated that they can adhere to medicine
- ▶ Clinical intermediaries have already been altered by the drug exposure

# Under-Ascertainment of Events That Occur Early in Therapy If Including Prevalent Users

- ▶ Classic example: the hormone replacement therapy story
- ▶ Worry that women using estrogen were at greater risk of heart attack/stroke
- ▶ Truth: increased risk of thromboembolic disease from exposure to estrogen/progestins *during first year of therapy (not later)*
- ▶ If we included only PREVALENT users, we would miss those early events
- ▶ Survivor bias—the women who stay on medicine past the first year have already “proven” that they tolerate the medicine

# Adherence Bias If Including Prevalent Users

- ▶ People who are adherent to medication have better outcomes (unrelated to the medicine)
- ▶ Individuals who are prevalent users of a drug have demonstrated that they can adhere (their experiences may then be different than new users whose adherence is unknown)

# Risk Factors May Be Altered by the Study Drugs If Including Prevalent Users

- ▶ With prevalent users, what are the *baseline* patient characteristics?
- ▶ Everyone is already on the drug, so lipids have been lowered ..., endometrium has thickened ...
- ▶ If these characteristics are intermediates to the clinical outcomes, they probably can't be adjusted for, as they are on the causal pathway

# Issues When Comparing New Drugs With Existing Therapies

- ▶ This has particular relevance to pharmacoepidemiology studies in which newly approved drugs are compared to existing therapies
- ▶ Possible explanation for the reported elevated risk of venous thromboembolic disease with third-generation progestins in oral contraceptives (OCPs)
- ▶ *Incident* users of the third-generation drugs were compared to *prevalent* users of established OCPs (not good design)

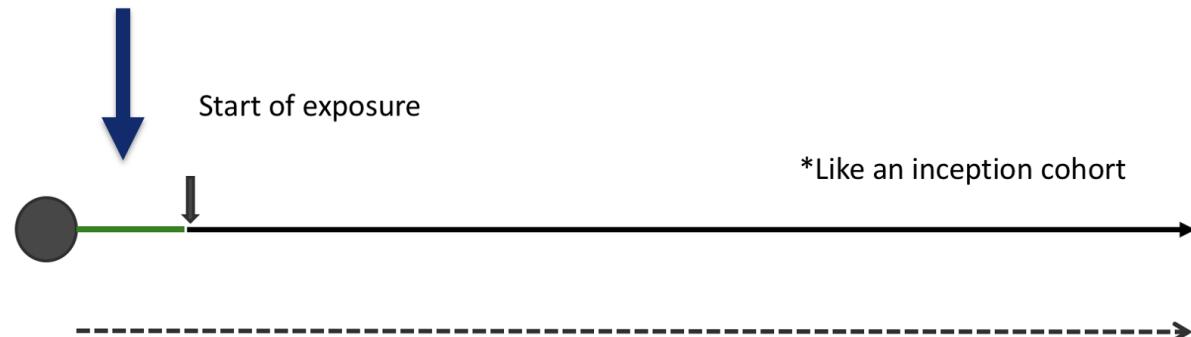
# New-Users

- ▶ Drug-exposed individuals are identified
- ▶ Individuals are required to have an observed period **free of exposure** before the exposure to be considered “new-users”
- ▶ Patient characteristics are assessed just before start of exposure



# New-User Design

- As in a clinical trial, the baseline measures are made before randomization and exposure to the intervention
- Baseline values before exposure can then be adjusted for if they are thought to be confounders

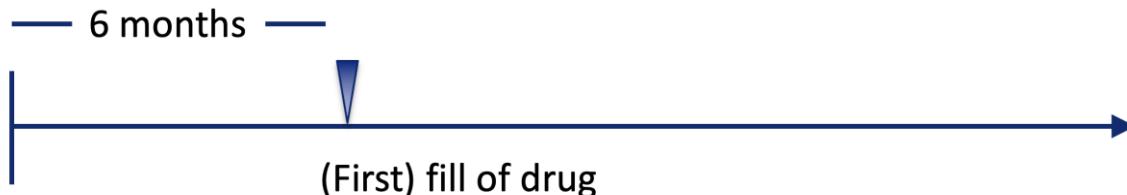


# Example of New-User Design

- ▶ **Study aim:** to evaluate the risk of suicidal acts or violent death associated with anticonvulsants
- ▶ **Data:** claims data from 2001 through 2006
- ▶ **Population:** individuals 15 years and older who **began taking** anticonvulsants between 2001 and 2006 with six months of enrollment **before** starting drug

# Design Choices

- ▶ Follow-up time began on the day after first exposure until drug stopped, drug switched, death, end of enrollment, end of data, or study outcome (suicide or violent death)
- ▶ People started on a new drug could contribute to the analyses for that drug if they had a six-month drug-free period before that drug started
- ▶ “Baseline” patient characteristics were assessed in first six months (demographics, comorbidities, other medication use with National Drug Codes)



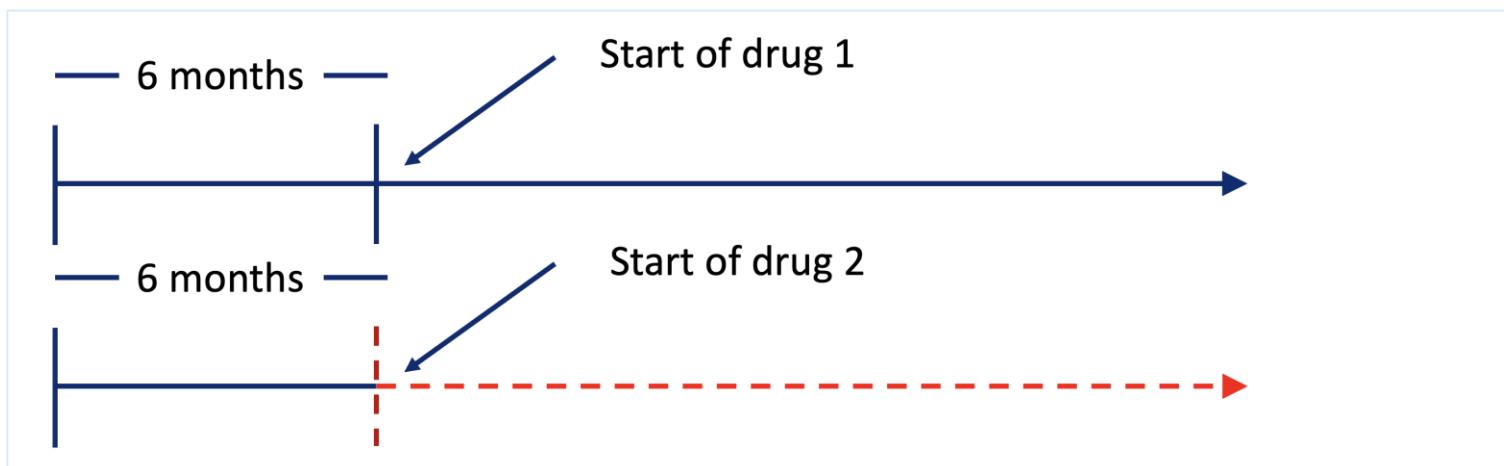
Source: Patorno, E., Bohn, R. L., Wahl, P. M., et al. (2010). Anticonvulsant medications and the risk of suicide, attempted suicide, or violent death. *JAMA*, 303(14), 1401–1409.

# Reducing Biases With Active Comparators

- ▶ Most pharmacoepidemiology research does not involve the comparison of drugs to placebos (we don't use placebos in clinical practice)
- ▶ The comparators are almost always “active comparators”—a different product for the same condition
  - ▶ Sometimes we compare a drug to a different drug
    - Compare selective serotonin reuptake inhibitors (SSRIs)
  - ▶ Sometimes we compare adding a drug to the initial drug alone
    - Add a dipeptidyl peptidase (DPP-4) to metformin compared to metformin alone
  - ▶ Sometimes we compare a drug to “usual care”
    - A mix of different exposures (a new antihypertensive compared to usual hypertension care)
- ▶ This reduces many misclassification biases

# Take-Home Message About Design

- When we design cohort studies in pharmacoepidemiology, they are most often:
- New-user, active comparator design**



# Immortal Time Bias

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

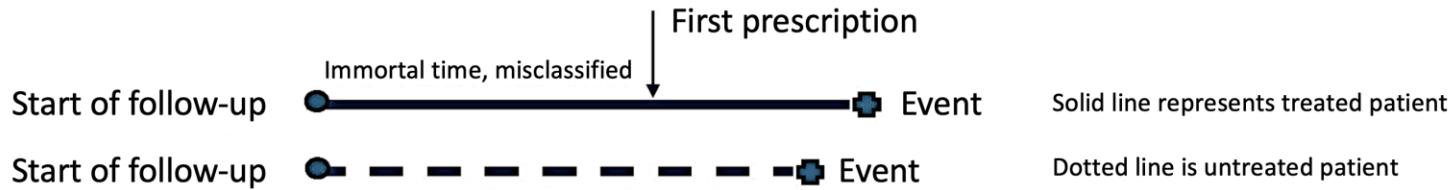
# Immortal Time (Bias) in Pharmacoepidemiology

- ▶ Immortal time is generated if there is a period of time during which an outcome event **cannot** occur (meaning that it is not counted)
- ▶ Immortal time typically arises when the determination of an individual's treatment status (exposure group) involves a waiting period
- ▶ Immortal time bias is somewhat unique to pharmacoepidemiology and needs to be avoided

# Be Careful of Any Waiting Periods

- ▶ This wait period is considered “immortal time” because individuals who end up in an exposure group had to survive (be alive and event free) until the treatment definition was fulfilled (or they wouldn’t be part of the exposed group)
- ▶ This biases the results in favor of the treatment under study by conferring a spurious survival advantage to the treated group

# Immortal Time Makes Misclassification Bias



- The period of immortal time is incorrectly attributed to the treated group if using a time-fixed analysis

Source: Lévesque, L. E., Hanley, J. A., Kezouh, A., & Suissa, S. (2010). Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ (Clinical Research Ed.)*, 340, b5087.

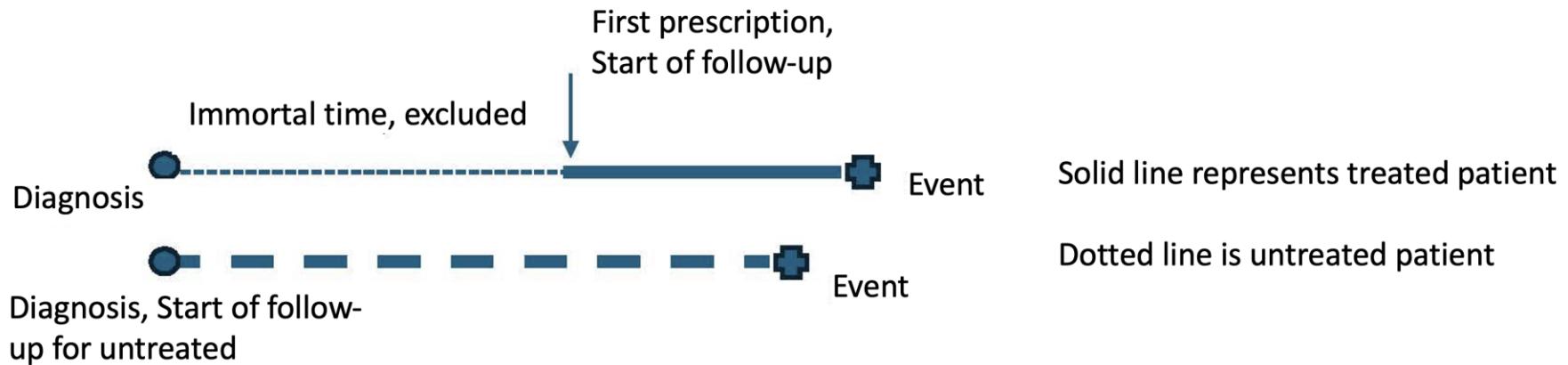
# Common Examples of Immortal Time (Don't Do This)

- ▶ Treatment defined as at least one prescription dispensed after hospital discharge, when the discharge date represents the start of follow-up (cohort entry)
  - ▶ For example, dispensation of an inhaled corticosteroid after a hospital stay for chronic obstructive pulmonary disease
- ▶ Treatment defined as at least one prescription dispensed after a diagnosis, when the date of diagnosis represents the start of follow-up
  - ▶ For example, starting interferon beta some time after diagnosis of multiple sclerosis

# Common Examples of Immortal Time (Seriously, Don't Do This)

- ▶ Treatment groups defined in terms of when after hospital discharge (start of follow-up) a prescription is dispensed
  - ▶ For example, cardiac drugs dispensed *within* seven days of discharge for acute myocardial infarction versus *later*, or early versus delayed dispensation of clopidogrel after a coronary intervention
- ▶ Treatment status determined over the duration of follow-up
  - ▶ For example, determining an individual's immunization status at the end of each influenza season or use of beta-blockers any time during follow-up

# Immortal Time: Selection Bias—Don't Do This

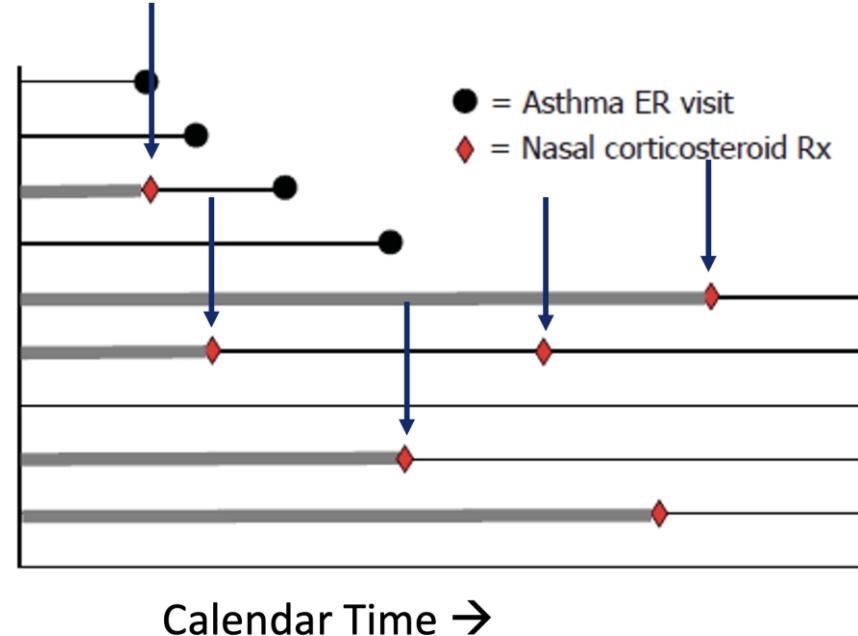


- The period of immortal time is *incorrectly* excluded from the analysis because the start of follow-up for the treated group is defined by the start of treatment and is, by design, later than that for the untreated group

Source: Lévesque, L. E., Hanley, J. A., Kezouh, A., & Suissa, S. (2010). Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ (Clinical Research Ed.)*, 340, b5087.

# Solutions

- ▶ A solution is to classify as unexposed the person-time before the prescription was received by using either a Poisson model approach or a Cox proportional-hazards model with time-dependent drug exposure
- ▶ Alternatively, redefining time zero as the day after the selected exposure; this removes the immortal time and then you can use a Cox model analysis without time-dependent exposures



# When Should You Worry That There Is Immortal Time in a Study?

- ▶ Criteria for identifying immortal time bias in cohort studies:
  - ▶ Was treatment status determined *after* the start of follow-up or defined using follow-up time?
  - ▶ Was the start of follow-up *different* for the treated and untreated (or comparator) groups relative to the date of diagnosis?
  - ▶ Were the treatment groups identified hierarchically (one group before the other)?
  - ▶ Were subjects excluded on the basis of a treatment that was identified *during* follow-up?
  - ▶ Was a time-fixed analysis used?

# Graphical Depiction of Study Design

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Graphical Depiction of Study Design: Overview

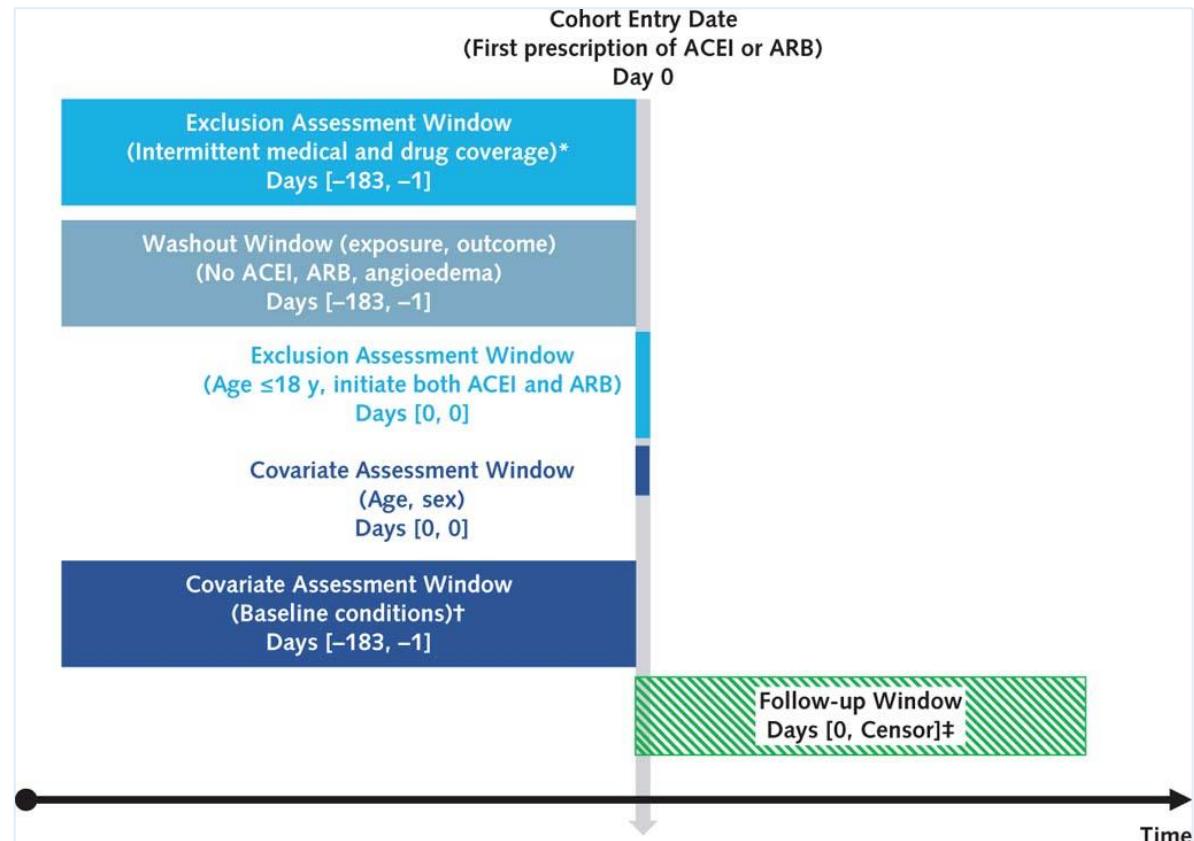
- ▶ As discussed, goal is to design a study with little bias
- ▶ A clear figure is extremely helpful when designing the study as you make decisions
- ▶ The periods of interest are the time intervals in which you look in the data for key events
- ▶ Nicely depicted in: Schneeweiss, S., Rassen, J. A., Brown, J. S., et al. (2019). Graphical Depiction of Longitudinal Study Designs in Health Care Databases. *Annals of Internal Medicine*, 170(6), 398–406.  
<https://doi.org/10.7326/M18-3079>

# Relevant Intervals (Example)

*ACEI = Angiotensin-converting enzyme inhibitor*

*ARB: Angiotensin II receptor blocker*

## A depiction of a cohort study design

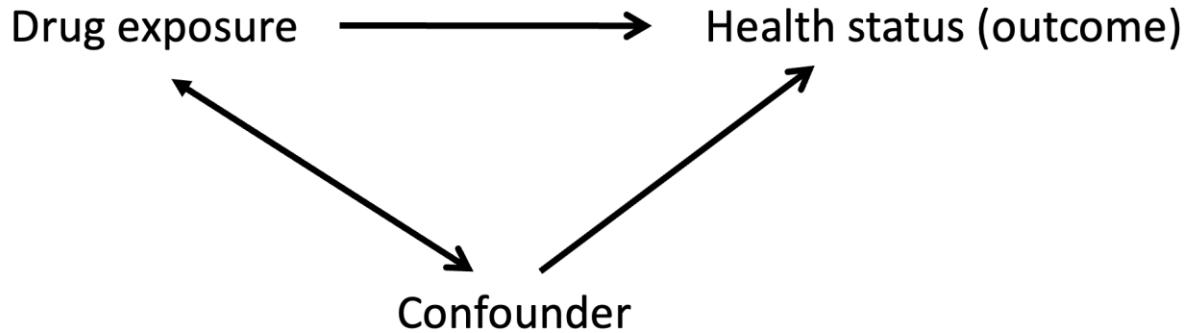


Source: Schneeweiss, S., Rassen, J. A., Brown, J. S., et al. (2019). Figure 1: Exposure-based cohort entry where the cohort entry date is selected before application of exclusion criteria [Diagram]. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. Annals of Internal Medicine, 170(6), 398–406. <https://doi.org/10.7326/M18-3079>

# Review of Confounding

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Review of Confounding Definition



- ▶ Measure of association between drug exposure and health status is distorted by the effect of one or several other variables that are also risk factors for the outcome, **and** these variables are imbalanced across exposure categories

# Confounding by Indication

- ▶ Most common type of confounding in pharmacoepidemiology:
  - ▶ “**Confounding by indication**”
  - ▶ “**Channeling**”
  - ▶ “**Indication bias**”
  - ▶ “**Confounding by severity**”
  - ▶ “**Contraindication bias**”

← These are all really the same issue

- ▶ There was a reason that the drug was prescribed, and that reason is associated with the outcome of interest

# Challenge of Confounding by Indication

- ▶ Closely related to selection bias
  - ▶ People are exposed or unexposed for some reason
  - ▶ We try to address selection bias at the time of study design
  - ▶ We try to address confounding by indication at the time of analysis (as best we can)
- ▶ Challenge is that it is hard to know the determinants of prescription of a medication (WHY)
  - ▶ If it was known that people with these characteristics are **always** exposed, this could be adjusted for, but this isn't the case

# Methods for Addressing Confounding by Indication

- ▶ Same options that are ordinarily used for confounding control in cohort studies
- ▶ Options
  - ▶ Matching
  - ▶ Subclassification or stratification
  - ▶ Adjustment (such as with regression models)
- ▶ Sometimes useful for these activities are:
  - ▶ Propensity scores
  - ▶ Disease risk scores (will not discuss here)
  - ▶ Instrumental variables (will not discuss here)



# Regression With Propensity Score Methods

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.

The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Propensity Score Methods

- ▶ Use of propensity score methods have become common in pharmacoepidemiology
- ▶ Not always necessary but often convenient and effective

# Why Might You Use Propensity Score Methods?

- ▶ In most studies with observational data, it is necessary to control for differences between the treated and untreated groups
  - ▶ These differences **would not** occur in a trial because of randomization
- ▶ Goal is to identify patients with identical likelihood of receiving treatment, but some did and some did not receive treatment (as if they were randomized!)
- ▶ A propensity score reduces the collection of background characteristics into a single composite characteristic, which can be useful

# Ordinarily, a Simple Confounding Problem

- ▶ Simple problem:
  - ▶ Age differs between treated and untreated individuals in this observational cohort

| Individuals        | Treated | Control |
|--------------------|---------|---------|
| Age (mean)         | 45      | 48      |
| Standard deviation | 6.3     | 7.2     |

- ▶ In analyses ...
  - ▶ We can adjust for age when estimating the treatment effect (regression)
  - ▶ Or stratify by age so that individuals <45 years and >45 years are analyzed separately
  - ▶ Or match individuals by age so that we have a sample with identical ages

# More Complicated

- ▶ Almost all covariates differ between treated and untreated individuals (each is only a small difference but contributes to a potentially large difference in outcomes that could confound the study)

|               | Treated  | Control  |
|---------------|----------|----------|
| Age (mean)    | 45 years | 46 years |
| Weight (mean) | 62 kg    | 66 kg    |
| Women         | 42%      | 44%      |
| Smokers       | 26%      | 27%      |
| Diabetes      | 32%      | 37%      |

- ▶ Could *adjust* for covariates if data set is big enough; could not really stratify by this many covariates, challenging to match on this many covariates

# Use of a Propensity Score

- ▶ Replace all of the baseline covariates with a single covariate that captures this information
- ▶ The covariate is a score that reflects the *propensity to receive the treatment of interest*
- ▶ It quantifies the probability that the patient was treated given his or her covariates

**$\text{Prob}(T = 1|X)$  where T is treatment and X are the covariates**

- ▶ Because of the balancing properties of the propensity score, samples or stratum that are **matched** on the propensity score will tend to be **balanced** in their distribution of observed covariates

# First Step

1. Estimate the propensity score for treatment as a function of observed covariates (covariates have to have been measured **before** the treatment)
  - ▶ Most use logistic regression
  - ▶ The clinical outcome variable is nowhere in the model

# Example

- ▶ We studied effect of exenatide on clinical outcomes (drug for treatment of diabetes) using administrative data
- ▶ Outcomes of interest were hospitalizations and costs
- ▶ We expected that the users of exenatide would be very different than the comparison group (new users of insulin)
- ▶ Why would they be different?
  - ▶ New drug, first-in-class, injectable, causes weight loss

# Example: Covariates Examined

*DM = Diabetes mellitus*

- ▶ Type 1 diabetes
- ▶ Duration of diabetes
- ▶ Concurrent medications
- ▶ Age
- ▶ Sex
- ▶ Obesity
- ▶ DM-related comorbidities
- ▶ Prescriber
- ▶ Many more

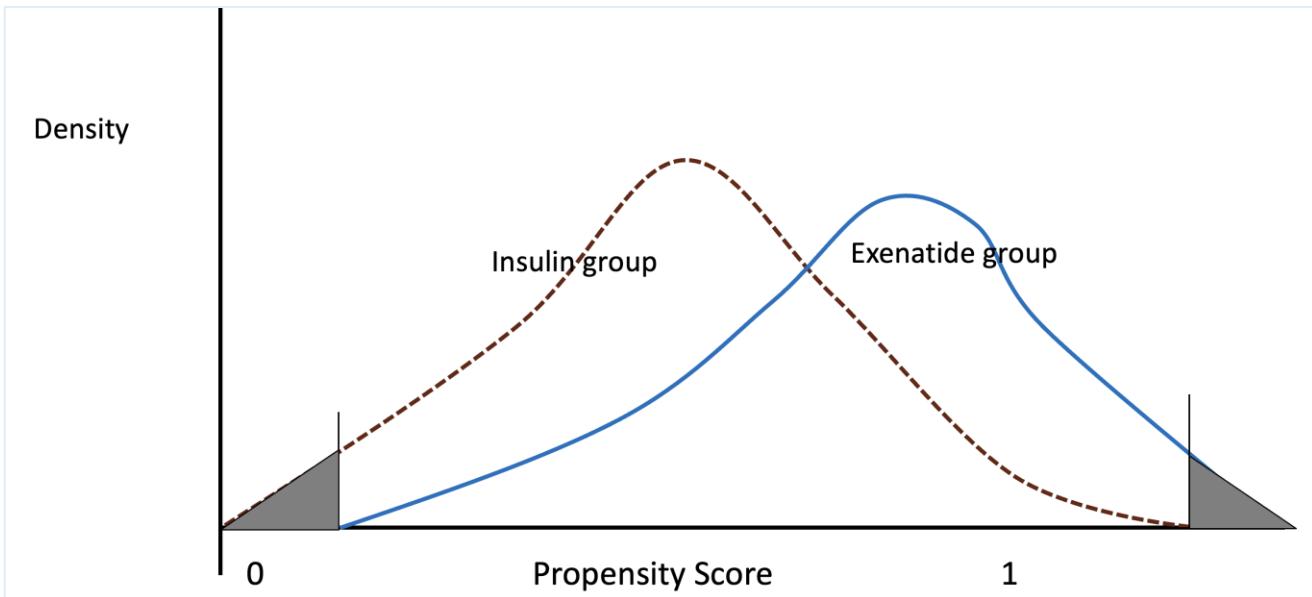


# Example: Generating the Score

- ▶ We generated a propensity score with logistic regression
  - ▶ Outcome variable was use of exenatide
  - ▶ Predictor variables were all the covariates that we thought might predict use of exenatide
- ▶ Logistic regression model predicted the probability that any individual would be prescribed exenatide (any individual has a probability from 0 to 1)

# Example of Propensity Score Distributions

Typical to “trim” the individuals with propensity scores in the non-overlapping regions. These are individuals with no chance of receiving that therapy (not really relevant to the study).



# Second Step

1. Estimate the propensity score for treatment as a function of observed covariates (covariates have to have been measured BEFORE the treatment)
  - ▶ May use logistic regression or discriminant analysis
  - ▶ Clinical outcome variable is nowhere in the model
2. Use the estimated propensity score to *control for confounding* in the models estimating treatment effect
  - ▶ Lots of choices

# What Can You Do With the Scores?

- ▶ You can sub-classify individuals (often into quintiles) based on their propensity scores and use this for adjustment in analyses
- ▶ Analyses are done as would be ordinarily done as with a stratified sample (e.g., Mantel-Haenzel odds ratios, or inclusion of strata as a categorical covariate in a logistic regression model)
- ▶ Stratum-specific treatment effects are valid to report

# What Else Can You Do?

- ▶ Can **match** on the propensity score—many options for making the matches
- ▶ Has the property of nicely balancing the characteristics that were included in the propensity score model
- ▶ Most studies report the standardized means differences in patient characteristics between the treatment groups after matching

# Example of Balance After Matching on Propensity Score

**Table 1. Baseline Characteristics of Patients Before and After Propensity Score Matching**

| Characteristic                            | Patient group             |                           |                         | After matching            |                        |                         |
|---|---------------------------|---------------------------|-------------------------|---------------------------|------------------------|-------------------------|
|   | Before matching           |                           | Standardized difference | After matching            |                        | Standardized difference |
|   | SGLT-2i use<br>(n = 5319) | Controls<br>(n = 225 047) |                         | SGLT-2i use<br>(n = 5317) | Controls<br>(n = 5317) |                         |
| <b>Demographic</b>                        |                           |                           |                         |                           |                        |                         |
| Age, mean (SD), y                         | 63.8 (12.3)               | 67.4 (15.5)               | 0.25                    | 63.8 (12.3)               | 64.2 (14.6)            | 0.03                    |
| Sex                                       |                           |                           |                         |                           |                        |                         |
| Men                                       | 3182 (59.8)               | 116 053 (51.6)            | 0.14                    | 3181 (59.8)               | 3175 (59.7)            | 0.002                   |
| Women                                     | 2137 (40.2)               | 108 994 (48.4)            | 0.14                    | 2136 (40.2)               | 2142 (40.3)            | 0.002                   |
| <b>Race and ethnicity</b>                 |                           |                           |                         |                           |                        |                         |
| American Indian or Alaska Native          | 29 (0.5)                  | 1009 (0.4)                | 0.01                    | 26 (0.4)                  | 30 (0.4)               | 0.01                    |
| Asian                                     | 141 (2.7)                 | 4518 (2.0)                | 0.04                    | 255 (4.8)                 | 255 (4.8)              | 0.002                   |
| Black                                     | 1008 (19.0)               | 40 782 (18.1)             | 0.02                    | 1122 (21.1)               | 1128 (21.2)            | 0.002                   |
| Native Hawaiian or Other Pacific Islander | 14 (0.3)                  | 369 (0.2)                 | 0.02                    | 70 (1.3)                  | 54 (1.0)               | 0.03                    |
| White                                     | 3495 (65.7)               | 146 890 (65.3)            | 0.02                    | 3493 (65.7)               | 3496 (65.8)            | 0.001                   |
| Other or unknown                          | 632 (11.9)                | 26 513 (11.8)             | 0.003                   | 710 (13.3)                | 704 (13.2)             | 0.001                   |
| <b>Comorbidities</b>                      |                           |                           |                         |                           |                        |                         |
| Hyperlipidemia                            | 3705 (69.7)               | 98 352 (43.7)             | 0.18                    | 3703 (69.6)               | 3254 (61.2)            | 0.18                    |
| Chronic kidney disease                    | 1806 (34.0)               | 63 527 (28.2)             | 0.11                    | 3533 (66.4)               | 3454 (65.0)            | 0.01                    |
| Hyperuricemia                             | 227 (4.3)                 | 264 (0.1)                 | 0.03                    | 227 (4.3)                 | 264 (5.0)              | 0.03                    |
| Congestive heart failure                  | 2716 (51.1)               | 57 320 (25.5)             | 0.53                    | 1806 (34.0)               | 1796 (33.8)            | 0.004                   |
| Ischemic heart diseases                   | 2969 (55.8)               | 77 556 (34.5)             | 0.42                    | 2967 (55.8)               | 2973 (55.9)            | 0.002                   |
| Cerebrovascular diseases                  | 1205 (22.7)               | 45 417 (20.2)             | 0.05                    | 1205 (22.7)               | 1187 (22.3)            | 0.01                    |
| Overweight                                | 2617 (49.2)               |                           | 0.42                    | 2615 (49.2)               | 2582 (48.6)            | 0.01                    |
| COPD                                      | 992 (18.7)                | 34 211 (15.2)             | 0.08                    | 991 (18.6)                | 961 (18.1)             | 0.02                    |
| Musculoskeletal disease                   | 3699 (69.5)               | 123 816 (55.0)            | 0.28                    | 3697 (69.5)               | 3704 (69.7)            | 0.003                   |
| Malignant neoplasm                        | 138 (2.6)                 | 5837 (2.6)                | 0.004                   | 138 (2.6)                 | 144 (2.7)              | 0.01                    |

Source: Pan, H. C., Chen, J. Y., Chen, H. Y., et al. (2024). *Table 1: Baseline characteristics of patients before and after propensity score matching [Table]. Sodium-Glucose Cotransport Protein 2 Inhibitors in Patients With Type 2 Diabetes and Acute Kidney Disease. JAMA Network Open, 7(1), e2350050. <https://doi.org/10.1001/jamanetworkopen.2023.50050> (boxes added)*

# Other Way to Use a Propensity Score?

- ▶ Can weight on the inverse of the propensity score (IPTW = inverse probability of treatment weighting)
- ▶ Requires decisions about the estimand
  - ▶ Are you estimating the average treatment effect for the treated (ATT)?
  - ▶ Are you estimating the average treatment effect for the whole population (ATE)?
- ▶ Each individual is weighted when included in the treatment effect model to estimate the risk (or odds or hazard) of the outcome

# What Variables Should Go Into the Propensity Score Model?

- ▶ Many opinions on this
- ▶ Studies using simulated data provide some guidance
- ▶ Some argue strongly that clinical knowledge should guide selection of variables for the model so that only true confounders are included
- ▶ Others argue that “high-density” propensity scores including many hundreds of variables and machine learning methods for selection of variables is more appropriate

# Limitations

- ▶ Propensity scores can adjust only for measured confounders
- ▶ Typically, this methodology should be used only with rather large samples
  - ▶ Otherwise, the covariates will still not end up balanced
- ▶ There is a potential problem if the treatment *effect* varies with the propensity score (which it often does)

# Effectiveness Trials

Produced by the Center for Teaching and Learning at the Johns Hopkins Bloomberg School of Public Health.  
The material in this video is subject to the copyright of the owners of the material and is being provided for educational purposes under rules of fair use for registered students in this course only. No additional copies of the copyrighted work may be made or distributed.

# Effectiveness Trials: Overview

- ▶ Pharmacoepidemiologists are often studying the comparative effectiveness and safety of approved products
- ▶ Goal of comparative effectiveness research is to generate better information about the risks and benefits and costs of different treatment options—which could eventually alter the way in which medicine is practiced and yield lower health care spending without having adverse effects on health
- ▶ Sometimes, we need to use randomization (effectiveness trials) to sufficiently answer these questions

# Pragmatic Clinical Trials: Definitions—1

“Should one prefer the goal of immediate applicability with a sacrifice of true understanding, or the more distant goal which may lead to greater enlightenment and which may prove more fertile for the future?”

—Schwartz, D., & Lellouch, J. (1967).

# Pragmatic Clinical Trials: Definitions—2

- ▶ Is the goal to acquire information about the true effects of a treatment, or is the goal to gather information needed to make a decision about a treatment?
- ▶ “Explanatory” trial versus “pragmatic trial”
- ▶ The context in which a trial is done determines if the trial is more explanatory or more pragmatic
- ▶ Used the terms: “laboratory,” which is associated with explanatory trials, and “normal,” which is associated with pragmatic trials



# Pragmatic Clinical Trials: Definitions—3

- ▶ Selection of outcomes to be evaluated in trials
  - ▶ If outcome is relevant to patients, evaluate in a pragmatic trial
  - ▶ If outcome has little biological information, evaluate in a pragmatic trial
- ▶ They used as an example the outcome of “returning to work”
- ▶ Explanatory approach: a strict patient selection criterion may be used so the population is homogenous
- ▶ Pragmatic approach: a heterogeneous population is acceptable

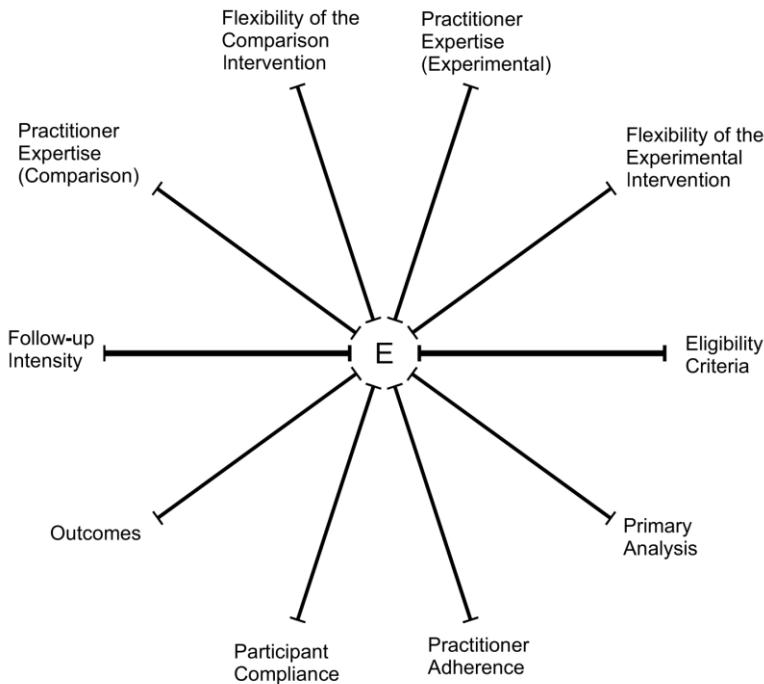
# Efficacy Versus Effectiveness Trials

|              | <b>Efficacy studies</b>  | <b>Effectiveness studies</b>  |
|--------------|--|---|
| Objective    | Does it work under optimal circumstances?                        | Does it work under usual circumstances?   |
| Motivation   | Regulatory approval—US Food and Drug Administration (FDA)        | Formulary approval  |
| Intervention | Fixed regimen/forced titration                                   | Flexible regimen  |
| Comparator   | Placebo; arbitrarily chosen comparator                           | Usual care, least expensive/most efficacious  |
| Design       | Randomized controlled trial (RCT) strict control                 | RCT or open label-minimum control   |
| Subjects     | Selected or “eligible” subjects; high compliance                 | Any subjects; low compliance  |
| Outcomes     | Condition-specific; strong link to mechanism; short-term horizon | Comprehensive (for example, quality of life, utilities); weak link to mechanism of action; short-term and long-term horizon |

# Pragmatic-Explanatory Continuum Indicator Summary (PRECIS)

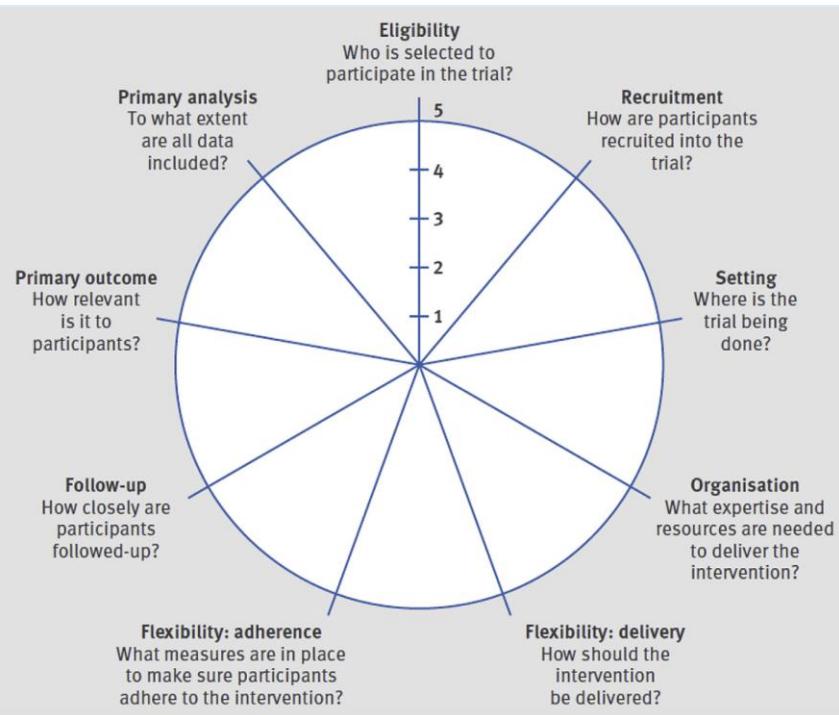
- ▶ Next advance grew out of discussion among investigators involved in the Pragmatic RAndomized Controlled Trials in Health Care (PRACTiHC) project, a Canadian and European Union initiative to promote pragmatic trials in low- and middle-income countries
- ▶ PRECIS: their definitions
  - ▶ Explanatory trial answers: “Can an intervention work under ideal conditions?”
  - ▶ Pragmatic trial answers: “Does the intervention work under usual conditions?”
- ▶ They developed the PRECIS Wheel—a graphic depiction of the decisions one makes when designing a trial

# Domains for the PRECIS Graphic



- The eligibility criteria for trial participants
- The flexibility with which the experimental intervention is applied
- The degree of practitioner expertise in applying and monitoring the experimental intervention
- The flexibility with which the comparison intervention is applied
- The degree of practitioner expertise in applying and monitoring the comparison intervention
- The intensity of follow-up of trial participants
- The nature of the trial's primary outcome
- The intensity of measuring participants' compliance with the prescribed interventions, and whether compliance-improving strategies are employed
- The intensity of measuring practitioners' adherence to the study protocol and whether adherence-improving strategies are employed
- The specification/scope of the analysis of the primary outcome

# PRECIS-2 for Decisions About Pragmatic Trial Design



- ▶ A scale for all domains which can be scored from 1 (very explanatory) to 5 (very pragmatic)
- ▶ Eligibility criteria has been placed on the 12 o'clock spoke
- ▶ Clockwise use of PRECIS-2 with logical ordering of adjacent domains
- ▶ The participant eligibility criterion has been separated into "Eligibility" and "Setting"
- ▶ New domains of "Recruitment" and "Organization"
- ▶ Domain name changes
- ▶ Removal of the comparison intervention domains of "Practitioner expertise" and "Flexibility of the comparison intervention"
- ▶ Domain name labels on the PRECIS-2 wheel now come with short explanations

# One Example

- ▶ Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., et al. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *The New England Journal of Medicine*, 353(12), 1209–1223.
- ▶ This trial is from the Clinical Antipsychotic Trials of Intervention Effectiveness Investigators (CATIE)
- ▶ To determine the comparative effectiveness of a representative antipsychotic (perphenazine) and several atypical antipsychotic medications for a representative sample of patients seeking treatment for chronic schizophrenia
- ▶ The primary outcome was time to all-cause treatment failure, marked by the discontinuation of the study medication
- ▶ For practice, consider reviewing this article and seeing what decisions the authors seem to have made using the PRECIS-2 wheel

# Study

- ▶ Few exclusion criteria (safety and treatment refractory individuals)
  - ▶ Patients and clinicians were masked to the treatment drug
  - ▶ Aimed to enroll 1,500 individuals from 50 clinical sites and follow them for 18 months
  - ▶ The treating physicians were allowed to titrate the medications to effectiveness
  - ▶ All patient participants were offered psychosocial interventions as well as an educational plan
- 
- ▶ For practice, use the PRECIS-2 wheel and review this article: did the study designers make pragmatic choices?

# Summary

- ▶ We reviewed some common biases including selection bias, information bias including misclassification bias, and immortal time bias
- ▶ We discussed how many biases can be overcome with use of a new-user, active comparator design
- ▶ We looked at a helpful graphical depiction of a cohort study
- ▶ We reviewed the concept of confounding and looked at use of propensity score methodology for matching, stratification, or regression as a way to control for confounders
- ▶ We reviewed key concepts about pragmatic trials which are sometimes needed for the study of effectiveness of drugs as used in practice