- In *Complications*, Atul Gawande, a famous surgeon-writer writes about his "great improbable save:"
- A 23-year-old woman comes to the ER with a local infection on the foot that looks everything like ordinary cellulitis, but which turns out to be necrotizing fasciitis
- Although there were no objective signs indicating necrotizing fasciitis, he makes this diagnosis early in the course of her illness partly (only?) because he happened to have seen another case of it recently
- There are no objective signs of NF, yet his gut feeling helps him make the diagnosis. How?
- What is gut feeling? Can we measure and quantify gut feeling? And how are diagnoses really made?

# Indication bias

Gustaf Edgren, MD PhD
Associate professor of epidemiology
Karolinska Institutet

## Outline

- Introduction
- When does it occur – examples
- How do we manage it?
  - Extending the epi toolbox
- What can we do when we can't manage it?

## What is indication bias?

- Indication bias is also referred to as confounding by indication, confounding by severity of disease, etc.
- *Probably the most important bias to keep in mind in clinical epidemiology*
- Results from the conscious choice of different treatments for patients with different prognosis

## What is indication bias? (2)

- In the classic case, it occurs in observational studies where

## Example 1 – Surviving the Holidays

- Hospital mortality is increased around major holidays in several settings
- We conducted a cohort study comparing 30-day mortality of patients transfused on major holidays
- Standard Cox regression adjusted for age, sex, number of transfusions, etc.
- Possible problems with this design?

## Surviving the holidays – results…

| Timing | Hazard ratio (95% CI) |
|---|---|
| New years eve | 1.22 (1.09-1.38) |
| Christmas eve | 1.15 (1.01-1.30) |
| Midsummer night | 1.04 (0.87-1.24) |
| Other weekends | 1.00 (ref) |
| Other weekdays | 0.74 (0.73-0.75) |

## Selection bias, *revisited*

- Selection bias occurs when the selection of subjects into a study is different depending on the exposure or disease status
  - In a cohort study, selection bias occurs when the selection of exposed and unexposed is related to the (future) disease status

## Confounding, *revisited*

- The effect of the exposure under study is mixed together with the effect of another variable. A confounder is:
  - associated with the disease
  - associated with the exposure
  - not an effect of the exposure

## Indication bias – Definition

- A type of selection bias (or, often, an example of confounding)
- In an external comparison (for example patients treated for condition X compared to the general population), it occurs when the condition for which the patients were treated is related to the outcome of interest

  *(This is, in my book, a typical example of selection bias)*

## Indication bias – Definition (2)

- Similarly, in an internal comparison (e.g. patients given drug A compared to patients given drug B), it occurs when the choice of drug is consciously or sub-consciously related to the risk of outcome of interest

  *(And this is a typical example of confounding…)*
- Is this common in observational studies?
- What about in RCT:s?

## What's in our toolbox?

- We've covered numerous methods for handling confounding and also how to avoid selection bias:
  - Stratification
  - Randomization
  - Adjustment
  - Etc.
- So, why another lecture on this?

## Rationale!

- Confounding by indication is probably…

THE MOST COMMON ERROR IN CLINICAL EPIDEMIOLOGY

- You need to learn how to manage or avoid it, and when that's not possible, to assess its possible effects

## Rationale! (2)

- So why is it such an important and difficult error?

ANYONE?

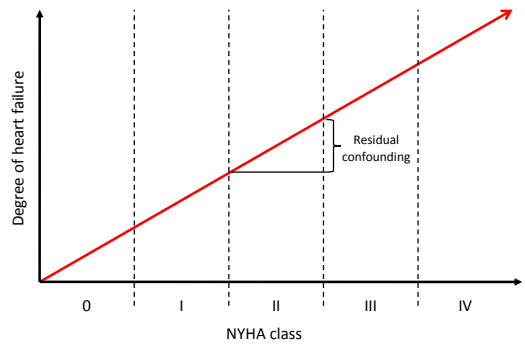
## Extended toolkit!

- Disease severity scores
- Propensity score method
- *Randomization – often with stratification*
- Natural randomization

## Disease severity scores

- For most diseases, clinicians and researchers have developed scoring systems for the classification of disease severity and level of symptoms
- Examples:
  - NYHA-classification for heart disease
  - TNM-classification of tumor spread
  - ISS (injury severity score) for trauma
  - Etc.

## Disease severity scores – problems?




## Disease severity scores (3)

- Generally, any attempt to classify severity of disease with a simple, one-dimensional score is likely to fail through lack of resolution
- That said, there are situations where we can still use scoring systems…
- For example, if nothing happens when you adjust for an imperfect score, then you're probably fine. But…

## Disease severity scores (4)

- Let's say you are doing a trauma study and you are estimating the relative risk of death in relation to the size of the hospital
- A likely confounder is severity of trauma (as measured with ISS), and so we adjust for this
- If the RR comparing small center to large center goes from 4 to 2 when we adjust for ISS, are we still fine?

## Disease severity scores – summary

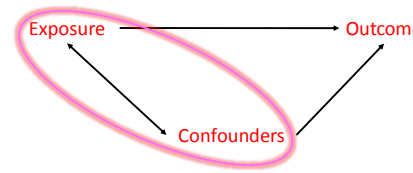
- Because of clinical nuances, there are relatively few situations where mere adjustment for severity scores succeeds in eliminating confounding by severity of disease
- It can, however, be used to estimate the direction and magnitude of the bias
- In some situations it can also indicate that such bias is NOT a big problem*

## Propensity score method

- The difficulties of adjusting for one (or many) severity scores often becomes forbidding in complex situations*
- One way of handling this is the propensity score method, according to which each subject is given a propensity score:
  - = P(treatment|covariate combination)
  - – (i.e., the conditional probability of receiving a certain treatment given a number of measured covariates
  - – Reduces all confounders to a single composite summary "score"

## Propensity score method

- So, why does a propensity score help us?
- The propensity score helps us compare subjects with different treatments who had the same propensity for treatment prior to the treatment decision




- Which helps minimize the confounding effects!

## PSS – how is it done?

- Usually, the PSS method is applied to clinical cohort (or C-C) studies with lots of clinical information
- The lazy epidemiologist then realizes that we only need to estimate P(treatment|covariates)
- This is usually done using logistic regression using which all subjects are allocated a propensity score: $0 \leq PS \leq 1$
- Importantly, the scoring is done with no heed to any outcome data. Exposure-outcome associations are estimated in a separate step!

## PSS – what covariates to include?

- The lazy epidemiologist might want to throw everything in there, but is that really the right way?
  - – Nay, if YOU (as a subject matter expert) do not believe that a factor is a possible confounder: IGNORE IT!
  - – As always, the covariates should also not be too strongly related to each other, and should not be too many, but otherwise common sense suffices

## PSS – what do we do with the score?

- Now, in the first part of a PSS analysis, we only assign the scores
- The scores themselves (that describe the association between treatment allocation and the combined effect of confounders) can be used for:
  - Adjustment (regression, stratification, etc.)
  - Matching in a case-control study
  - Stratification in an RCT

## Propensity score method – limitations

- Beyond where treatment allocation rests on unmeasureable nuances – a situation where no score will make a large difference – the propensity score method is quite robust
- However, as always, do not expect to be able to avoid all effects of indication bias
- Also, there must be an overlap in score between treatment groups!
- For some reason (conservatism?) propensity scores are quite seldom used, which may be a problem when publishing your brilliant study

## Natural (?) randomization

- When observational studies of treatment effects fail, researchers often claim that randomization is the only solution
- However, pessimistically, randomization is rarely feasible and almost never practical
- Thus, we need other solutions…

## Natural randomization (2)

- In some situations, treatment follows guidelines that are very strict and give little opportunity for an association with disease severity
- If, in such situations, guidelines differ between hospitals or clinics, treatment can be considered pseudo randomized
- Naturally, this is far from the validity of an RCT, but may be suitable for homogenous diseases

## Natural randomization (3)

- Other examples of natural randomization have been covered earlier in the course:


Risk of cancer after blood transfusion from donors with subclinical cancer: a retrospective cohort study

Gustaf Edgren*, Henrik Hjalgrim*, Marie Reilly*, Trung Nam Tran, Klaus Rostgaard, Agneta Shanwell, Kjell Titlestad, Johanna Adami, Agneta Wikman, Casper Jersild, Gloria Gridley, Louise Wideroff, Olof Nyrén, Mads Melbye

## True randomization

- We've already covered randomization!
- While randomization is a conceptually simple way of handling indication bias, it still pays to keep these matters in mind when you do randomize
- For example, if disease severity is an important predictor – stratify for it using a disease severity score!

## Indication bias – summary

- Indication bias (confounding by indication/severity of disease) is one of the most important biases to keep in mind in clinical epidemiology
- Researchers often attempt to adjust this away, but due to the complex decision making in medicine, they are often unsuccessful
- At our disposal, we have a variety of tools

## Indication bias – summary (2)

- Disease severity scoring systems are usually one-dimensional, and rather simplistic tools that do measure disease severity, but are limited by residual confounding
- Propensity scores are used to create a summary measure of the association between *several* covariates and treatment allocation
- Ultimately, confounding by severity can only be completely avoided by randomization