# Lessons about Confounding and Selection Bias Taught by the New User Design

Til Stürmer, FISPE

UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill, USA

## Conflict of Interest

- I receive investigator initiated funding (R01 AG023178) from the National Institute on Aging of the NIH
- I receive research funding as PI of the UNC-DEcIDE Center of the US Agency for Health Care Research and Quality (AHRQ)
- I receive salary support as the director of the UNC Center of Excellence in Pharmacoepidemiology and Public Health
- I receive salary support from unrestricted research grants from pharmaceutical companies (Merck, Sanofi-Aventis) to UNC
- I do not accept personal compensation of any kind from any pharmaceutical company

## Outline

- Why we need nonexperimental studies of medical interventions (brief)
- Confounding
  - By indication
  - By frailty
  - Study design options to limit confounding
  - Study design options to address unmeasured confounding
- Selection bias
  - By conditioning on persistence
  - Study design options to limit bias

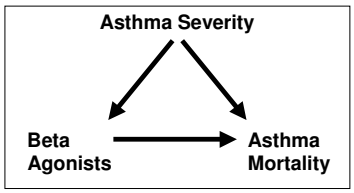
## Nonexperimental Studies of Medical Interventions

- Needed due to shortcomings of RCTs
  - Population, comorbidity, comedication, N, timeliness
  - E.g., benefit/harm of chemotherapy in > 75 year old
- Large population based linked healthcare databases
  - Solve all of the above
  - But: intractable confounding by indication for benefits (Miettinen 1983; Yusuf, Collins, & Peto 1984)
  - Sicker patients more likely to be treated

## Confounding by Indication

- Good prescribing leads to confounding of drug effects on intended outcomes
- More severe disease more likely to
  - Be treated (with higher doses)
  - Have higher risk of adverse outcomes
- Assessment of severity of disease
  - Often difficult
  - Sometimes impossible
- Drug looks BAD compared with **NON-USERS**!

## (Arche)typical Example for Confounding by Indication




Note: Problem is that we cannot measure asthma severity well (not the confounding per se)!

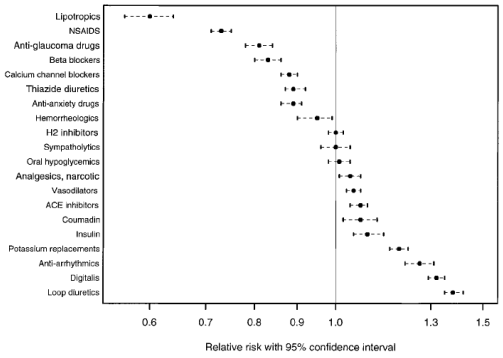
## Confounding by Frailty in Population Based PE Studies

- Individuals close to death are
  - Less likely to receive preventive treatments
    - E.g., statins, flu vaccination
  - More likely to receive palliative treatments
    - E.g., opiates instead of NSAIDs
  - More likely to receive certain classes of drugs
    - E.g., loop diuretics vs. other diuretics
- Paradoxical drug mortality associations
- Drug looks GOOD vs. **NON-USERS!**

## Frailty

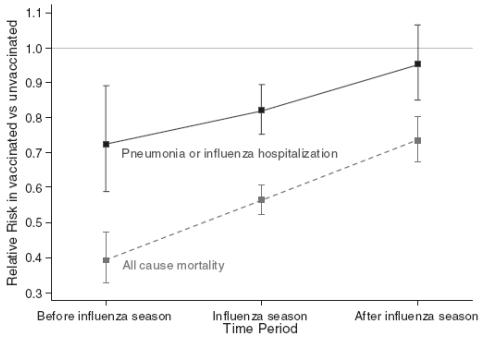
- End of life loss of
  - Weight
  - Physical function
  - Cognitive function
- Recognized by healthcare professionals
- Reduces likelihood of preventive therapies
  - Focus on main medical problem (Redelmeier et al. 98)
  - Little expected benefit (competing risks; Welch et al. 96)

## (Arche)typical Example for Confounding by Frailty




RR for 1-year mortality, hospitalized Medicare population.
Glynn RJ et al. Paradoxical Drug Effects in Elderly. Epidemiology 2001

## (Arche)typical Example for Confounding by Frailty




RR for mortality in >65-year old HMO population
Jackson LA et al. Int J Epidemiol 2006

## Methods to Limit Confounding by Indication and Frailty

- Compare initiators of treatments that are interchangeable within the same indication
- Exclude patients without “equipoise” based on propensity scores/instrumental variables
- Exchangeable cohorts of new users
- Essentially eliminates confounding by frailty
- Reduction of confounding by indication needs to be assessed for each comparator

## Purely Hypothetical Studies

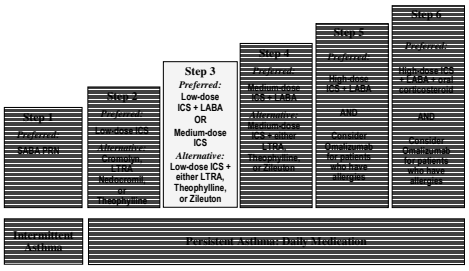
- Insulin analogues and risk for cancer
  - Confounding by unmeasured BMI
- Long-acting beta-agonists and risk for asthma mortality
  - Residual confounding by asthma severity
  - Archetype of confounding by indication

## Nonexperimental Study Comparing New Users of Insulin Analogue with Human Insulin

| BMI category | Insulin A | Insulin B |
|---|---|---|
| <19 | 7 ( 0.9%) | 14 ( 2.3%) |
| 19-<25 | 103 (13.1%) | 115 (18.8%) |
| 25-<30 | 205 (26.0%) | 151 (24.7%) |
| 30-<35 | 210 (26.6%) | 156 (25.5%) |
| 35-<40 | 148 (18.8%) | 84 (13.8%) |
| 40-<45 | 59 ( 7.5%) | 48 ( 7.9%) |
| ≥ 45 | 47 ( 6.0%) | 39 ( 6.4%) |
| missing | 10 ( 1.3%) | 4 ( 0.7%) |

Note 1: this is NOT a randomized study!
Note 2: this will not ALWAYS work!

## Nonexperimental Study Comparing New Users of LABA with ?




**National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program; Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma; Full Report 2007**

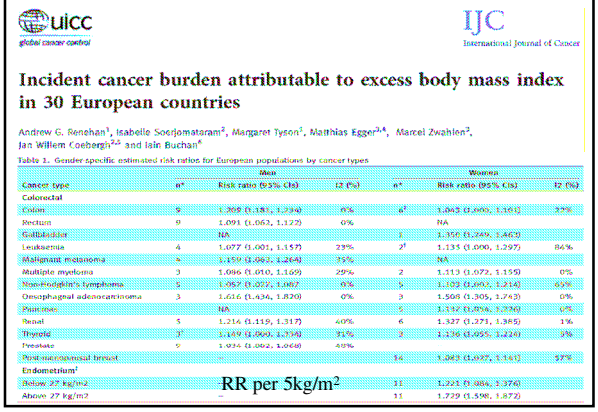
## Limit Potential for Confounding by Study Design

- New user design (Kramer 87, Miettinen & Caro 89, Guess 89, Moride & Abenhaim 94, Ray 03)
- At a point in time where something happens
- Comparator drug (CER)
  - Therapeutic alternative
  - Same stage of disease progression
  - Same (similar) indication
- Confounding limited to choice between alternatives
- Ideally for same indication
  - Insulin analogue vs. NPH insulin: lack of control oral hypoglycemics
  - LABA vs. ? added to ICS: lack of control under ICS alone
  - Note: guidelines and expert knowledge needed here!

## Additional Steps to Address Unmeasured Confounding

- Effect of unmeasured confounder on outcome of interest often well described
  - Reason to bother about BMI, asthma severity!
- Effect of unmeasured confounder on treatment decisions often unknown
  - Especially given measured covariates
  - May be estimated in external validation study using propensity scores (BMI data presented from EMR study)

## What if BMI Not Balanced?



uicc global cancer control

IJC International Journal of Cancer

**Incident cancer burden attributable to excess body mass index in 30 European countries**

Andrew G. Renehan[1], Isabelle Soerjomataram[2], Margaret Tyson[2], Matthias Egger[3,4], Marcel Zwahlen[3], Jan Willem Coebergh[2,5] and Iain Buchan[6]

Table 1. Gender specific estimated risk ratios for European populations by cancer types

| | Men | | | Women | | |
|---|---|---|---|---|---|---|
| Cancer type | n* | Risk ratio (95% CIs) | I2 (%) | n* | Risk ratio (95% CIs) | I2 (%) |
| Colorectal | | | | | | |
| Colon | 9 | 1.308 (1.181, 1.734) | 0% | 6† | 1.043 (1.000, 1.161) | 37% |
| Rectum | 9 | 1.091 (1.063, 1.122) | 0% | | NA | |
| Gallbladder | | NA | | 1 | 1.355 (1.249, 1.469) | |
| Leukaemia | 4 | 1.077 (1.001, 1.157) | 23% | 2† | 1.135 (1.000, 1.297) | 86% |
| Malignant melanoma | 5 | 1.159 (1.083, 1.264) | 75% | | NA | |
| Multiple myeloma | 3 | 1.096 (1.010, 1.169) | 29% | 2 | 1.113 (1.072, 1.155) | 0% |
| Non-Hodgkin's lymphoma | 5 | 1.052 (1.022, 1.087) | 0% | 5 | 1.109 (1.003, 1.214) | 65% |
| Oesophagal adenocarcinoma | 3 | 1.616 (1.434, 1.820) | 0% | 3 | 1.506 (1.305, 1.743) | 0% |
| Pancreas | | NA | | 5 | 1.132 (1.004, 1.230) | 0% |
| Renal | 5 | 1.214 (1.119, 1.317) | 40% | 6 | 1.127 (1.071, 1.185) | 1% |
| Thyroid | 3† | 1.149 (1.000, 1.336) | 31% | 3 | 1.136 (1.055, 1.224) | 0% |
| Prostate | 9 | 1.094 (1.062, 1.040) | 48% | | | |
| Post-menopausal breast | | | | 14 | 1.083 (1.022, 1.161) | 37% |
| Endometrium‡ | | | | | | |
| Below 27 kg/m2 | | – | | 11 | 1.221 (1.086, 1.376) | |
| Above 27 kg/m2 | | – | | 11 | 1.729 (1.598, 1.872) | |

RR per 5kg/m$^2$

## External Adjustment for Confounding

- Estimate effect of unmeasured confounder on treatment decision in (external) validation study
- Get estimate of the effect of the unmeasured confounder on the outcome from the literature
- “Adjust” main study estimate for unmeasured confounding using standard formulas (Bross 66)
- Separate estimates of the above for multiple confounders
- Weighted average of expected confounding
- Magnitude and overall direction of confounding

## Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information

*The Example of COX2 Inhibitors and Myocardial Infarction*

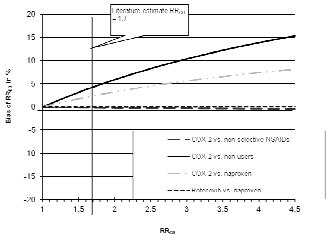
| | | High School Education or Less | Income | | Aspirin | Smoking Status | | |
|---|---|---|---|---|---|---|---|---|
| | No. | ≤$20,000 % | Obese* % | User % | Current % | Former % | Never % |
| Any COX2 inhibitor† | 872 | 69 | 48 | 24 | 9 | 8 | 45 | 46 |
| Celecoxib only | 562 | 69 | 46 | 24 | 8 | 9 | 45 | 46 |
| Rofecoxib only | 244 | 66 | 48 | 19 | 11 | 7 | 49 | 44 |
| Nonselective NSAIDs only | 1302 | 72 | 56 | 24 | 10 | 10 | 51 | 40 |
| Ibuprofen | 281 | 73 | 58 | 23 | 14 | 15 | 50 | 16 |
| Naproxen | 238 | 74 | 55 | 20 | 8 | 8 | 54 | 38 |
| Other NSAIDs only | 677 | 71 | 55 | 25 | 8 | 9 | 50 | 41 |
| Nonusers | 6611 | 69 | 53 | 17 | 9 | 10 | 50 | 40 |

Schneeweiss, Glynn, Tsai, Avorn, Solomon; Epidemiology 2005

## Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information

*The Example of COX2 Inhibitors and Myocardial Infarction*

BMI (obese vs. non-obese)




Schneeweiss, Glynn, Tsai, Avorn, Solomon; Epidemiology 2005

http://www.drugepi.org/wp-

## External Adjustment for Confounding

**Practice of Epidemiology**

**Propensity Score Calibration in the Absence of Surrogacy**

**Mark Lunt, Robert J. Glynn, Kenneth J. Rothman, Jerry Avorn, and Til Stürmer***

* Correspondence to Dr. Til Stürmer, Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, McGavran-Greenberg, CB 7435, Chapel Hill, NC 27599-7435 (e-mail: til.sturmer@post.harvard.edu).

*Initially submitted February 10, 2011; accepted for publication November 14, 2011.*

Propensity score calibration (PSC) can be used to adjust for unmeasured confounders using a cross-sectional validation study that lacks information on the disease outcome (*Y*), under a strong surrogacy assumption. Using directed acyclic graphs and path analysis, the authors developed a formula to predict the presence and magnitude of the bias of PSC in the simplest setting of a binary exposure (*T*) and 1 confounder (*X*) that are observed in the main study and 1 confounder (*C*) that is observed in the validation study only. PSC bias is predicted on the basis of parameters that can be estimated from the data and a single unidentifiable parameter, the relative risk (RR) associated with *C* ($RR_{CY}$). The authors simulated 1,000 cohort studies each with a Poisson-distributed outcome *Y*, varying parameter values over a wide range. When using the true parameter for $RR_{CY}$, the formula predicts PSC bias almost perfectly in this simple setting (correlation with observed bias over 24 scenarios assessed: $r = 0.998$). The authors conclude that the bias from PSC observed in certain scenarios can be estimated from the imbalance in *C* between treated and untreated persons, after adjustment for *X*, in the validation study and assuming a range of plausible values for the unidentifiable $RR_{CY}$.

## Sensitivity Analysis for Unmeasured Confounding

**Practice of Epidemiology**

**Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution—A Simulation Study**

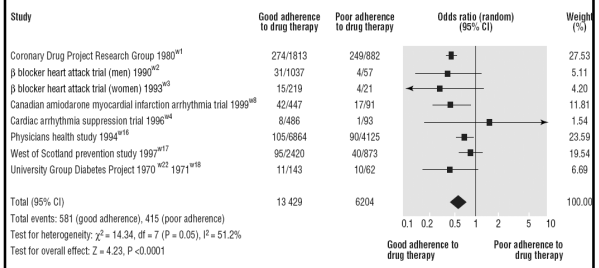
**Til Stürmer*, Kenneth J. Rothman, Jerry Avorn, and Robert J. Glynn**

* Correspondence to Dr. Til Stürmer, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, McGavran-Greenberg Hall, CB 7435, Chapel Hill, NC 27599-7435 (e-mail: til.sturmer@post.harvard.edu).

*Initially submitted February 23, 2010; accepted for publication May 26, 2010.*

Frailty, a poorly measured confounder in older patients, can promote treatment in some situations and discourage it in others. This can create unmeasured confounding and lead to nonuniform treatment effects over the propensity score (PS). The authors compared bias and mean squared error for various PS implementations under PS trimming, thereby excluding persons treated contrary to prediction. Cohort studies were simulated with a binary treatment *T* as a function of 8 covariates ***X***. Two of the covariates were assumed to be unmeasured strong risk factors for the outcome and present in persons treated contrary to prediction. The outcome *Y* was simulated as a Poisson function of *T* and all ***X***'s. In analyses based on measured covariates only, the range of PS's was trimmed asymmetrically according to the percentile of PS in treated patients at the lower end and in untreated patients at the upper end. PS trimming reduced bias due to unmeasured confounders and mean squared error in most scenarios assessed. Treatment effect estimates based on PS range restrictions do not correspond to a causal parameter but may be less biased by such unmeasured confounding. Increasing validity based on PS trimming may be a unique advantage of PS's over conventional outcome models.

## Selection Bias: Adherence to Placebo and Mortality




**OR for good vs. bad adherence and all cause mortality - Simpson, S. H et al. BMJ 2006**

## Selection Bias

- Better adherence associated with ~50% lower mortality (unless therapy harmful)
- Most likely explanation: sick stoppers
- Frail participants stop adhering to trial protocol
- Similar issue with persistence on preventive drugs?

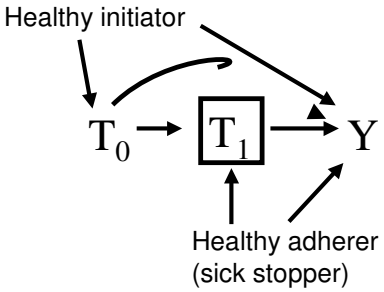
# Statin Persistence

**Table 3. Association Between Adherence to Statin Therapy and Risk of Health-Related Events**

| Outcome | More Adherent Event Rate, /100 Person-Years | Less Adherent Event Rate, /100 Person-Years | Unadjusted HR | 95% Confidence Limits for HR | Adjusted HR | 95% Confiden Limits for HF |
|---|---|---|---|---|---|---|
| Accident events | | | | | | |
| Both sexes (n=141 086) | | | | | | |
| Burn | 0.28 | 0.36 | 0.78 | (0.71–0.87) | 0.88 | (0.79–0.97) |
| Fall | 0.53 | 0.54 | 0.98 | (0.90–1.06) | 0.90 | (0.83–0.98) |
| Fracture | 2.20 | 2.38 | 0.93 | (0.89–0.96) | 0.92 | (0.88–0.96) |
| Motor vehicle accident | 1.48 | 2.25 | 0.66 | (0.63–0.69) | 0.75 | (0.72–0.79) |
| Open wound | 2.44 | 2.74 | 0.89 | (0.86–0.92) | 0.91 | (0.88–0.95) |
| Poisoning | 0.32 | 0.41 | 0.78 | (0.71–0.86) | 0.86 | (0.78–0.94) |
| Workplace accident | 1.31 | 2.13 | 0.62 | (0.59–0.65) | 0.77 | (0.74–0.81) |
| All (first occurrence) | 7.38 | 9.39 | 0.79 | (0.77–0.81) | 0.85 | (0.83–0.87) |
| Screening events | | | | | | |
| Both sexes (n=141 086) | | | | | | |
| Eye examination | 3.58 | 2.93 | 1.21 | (1.17–1.26) | 1.08 | (1.05–1.12) |
| Fecal occult blood test | 8.06 | 6.14 | 1.31 | (1.27–1.34) | 1.21 | (1.18–1.24) |
| Sigmoidoscopy | 0.53 | 0.49 | 1.09 | (1.00–1.18) | 1.07 | (0.98–1.16) |
| All (first occurrence) | 12.01 | 9.28 | 1.28 | (1.25–1.31) | 1.17 | (1.15–1.20) |

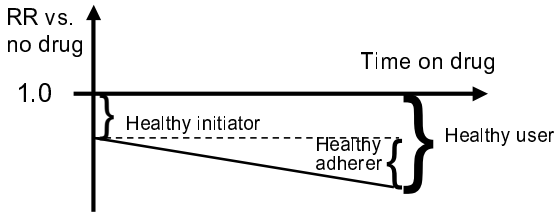
BC data; Dormuth et al. Circulation 2009

# Healthy User Bias DAG




Conditioning on $T_1$ due to censoring if $T_0$=1 & $T_1$=0

# Bias Over Time on Preventive Drug Compared with Non-Use



## Study Design Options to Limit Selection Bias

- Cannot be “controlled” for (introduced by control)!
- Cannot be “avoided” by excluding stoppers-switchers-augmenters from cohort (introduced by)!
- Solution: first treatment carried forward (ITT)
  - Standard for RCT (efficacy): tends to bias towards null
  - Works best for short term intended effects
- Always present in as treated analyses, but:
  - Limited by adding various lag times
    - Reduces bias from informative censoring (extreme: ITT)
  - Marginal structural models
    - Dependent on prediction of treatment(s) based on all risk factors for the outcome

## So How Does This All Fit Together?


## Confounding and Selection Bias in Population Based Studies of Medical Interventions

- Study design
  - New user design
  - Comparator drug
  - Sensitivity analyses on stopping/switching/augmenting
  - Consider internal or external validation studies to control for unmeasured confounders
  - Consider methods not based on no unmeasured confounding assumption (i.e., instrumental variables, self-controlled designs)
- Study analysis
  - Look for nonuniform effects e.g., over range of PS
    - Consider PS matching, range restrictions, trimming
    - Discuss unmeasured confounding vs. treatment heterogeneity

## Take Home Message

- Confounding by indication and frailty major threats to validity of nonexperimental studies
- Selection bias major threat to validity of any study (not specific to nonexperimental studies)
- Advances in PE study design allow us to reduce potential for bias
  - New user design and active comparator cohorts
  - Implement various valid alternatives to handle changes in treatment during follow-up
- Proper research question and study design more important than analysis