



# Dissecting a cohort study

David Madigan<sup>1</sup>

Patrick Ryan<sup>2</sup>

Martijn J. Schuemie<sup>2</sup>

Marc A. Suchard<sup>3</sup>

1. Department of Statistics, Columbia University
2. Janssen Research & Development, LLC
3. Departments of Biomathematics and Human Genetics, David Geffen School of Medicine, and Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles



# Exercise

Paper: Graham *et al.* (2015) Circulation

*“Cardiovascular, bleeding and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation”*

1. Team up into groups of 4 (2 programmer/designer pairs)
2. Discuss:
  - How to **reproduce** study
  - Shortcomings of design or analysis that limit **reliability**
3. Share with class

All in 60 minutes



# Getting started

Stated design: “*A new-user retrospective cohort design ...*”

1. What are the target (T) and comparator (C) cohorts?
2. How are these populations made *more* exchangeable?
3. How is the outcome (O) cohort defined?
4. What is the time-at-risk for the outcome?
5. How are outcome rate differences estimated and assessed?

**Important:** Look for ambiguity or lack of transparency

We want **reproducible** and **reliable** research!



# Example

1. What are the target (T) and comparator (C) cohorts?
  - How are the **medical concepts** in the inclusion/exclusion criteria defined?

Now, get to work!



# T and C cohorts

- Elderly ( $\geq 65$ ) Medicare beneficiaries (A, B and D) with nonvalvular atrial fibrillation who initiated therapy with dabigatran (T) or warfarin (C)

Is this correct?



# Inclusion criteria

All patients who:

- Have any inpatient or outpatient AF or atrial flutter ICD9 codes
- Filled at least 1 prescription for either drug between Oct 19, 2010 - Dec 31, 2012

Index date: first prescription date

- How are “inpatient or outpatient AF or atrial flutter ICD9 codes” defined?

Supplementary material provides codes for outcomes only.



# !Inclusion (exclusion) criteria

All patients who:

- Have > 6 months of Medicare enrollment before index date
- 
- Received prior treatment (when?) with NOAC or warfarin
- 
- Were in hospice on index date (why?)
- 
- Undergoing dialysis (when?)
- 
- Had diagnoses of valvular disease, DVT, PE, joint replacement during baseline 6 months



# !Inclusion (exclusion) criteria

All patients who:

- Have < 6 months of Medicare enrollment before index date
- Were < 65
- Received prior treatment (when?) with NOAC or warfarin
- Were in a skilled nursing facility on index date (why?)
- Were in hospice on index date (why?)
- Had a hospitalization “that extended beyond the index dispensing date”
- Undergoing dialysis (when?)
- Were kidney transplant recipients
- Had diagnoses of valvular disease, DVT, PE, joint replacement during baseline 6 months



# !Inclusion (exclusion) criteria

All patients who: (subset)

- Received prior treatment (when?) with NOAC or warfarin
- Had diagnoses of valvular disease, DVT, PE, joint replacement during baseline 6 months

Questions:

1. Prior treatment during baseline 6 months or anytime before index data?
2. How are the medical concepts defined?
3. What about outcomes prior to exposure?



# Balancing cohorts

Using propensity score model

- Logistic regression with “initiated dabigatran” as outcome and predictors:



# Balancing cohorts

## Using propensity score model

- Logistic regression with “initiated dabigatran” as outcome and predictors:

Sociodemographics (?)

Prescriber characteristics

Baseline comorbidities, medications (?)

“Other potentially relevant variables”

How were these chosen? Other diagnostics?

- 1:1 ratio, greedy matching
- Balance assessment via:
  - Standardized mean difference (target:  $\leq 0.1$ )

What was not balanced before/after matching?



# Outcome definitions

Risk of:

- Stroke
- Major gastrointestinal and intracranial bleeding
- Acute myocardial infraction
- Mortality

1. How were the outcomes defined?
2. Did the researchers get all of the appropriate codes?
3. Were each outcome analyzed independently or were the T+C cohorts constituted only once?
4. How does the proceeding influence the analysis?



# Changing outcomes

Overlapping authors have an alternative protocol under Mini-Sentienel.

Protocol

How and why do the outcome definitions differ?



# Outcome and time-at-risk

Time-to-event (first) event analysis using a **Cox proportional hazards model**

- Follow-up starts on index date + 1 and censored at:

Medicare disenrollment

> 3 day gap in anticoagulant supply

RX fill for a different anticoagulant

Start of hospice

Initiation of dialysis or kidney

transplant Admission to nursing

facility

End of study

How does exposure duration influence the effect estimate?



# Reliability

How do the authors assess the reliability of their estimates?

1. Should use **negative controls** to measure residual bias (*systematic error*) and ...



# Sensitivity to time-at-risk

Restricted to patient with:

- Initial RX  $\leq$  30 days
- $\geq$  2 RX fills
- Increased gap allowance to 14 days

How does exposure duration *now* influence the effect estimate?



# Subgroup analysis: age, sex

Subgroup analyses are *generally* a **bad** idea. Instead, use **interaction terms** in the outcome regression model.

- Reduces multiple testing trouble
- Helps adjust for correlation between predictors

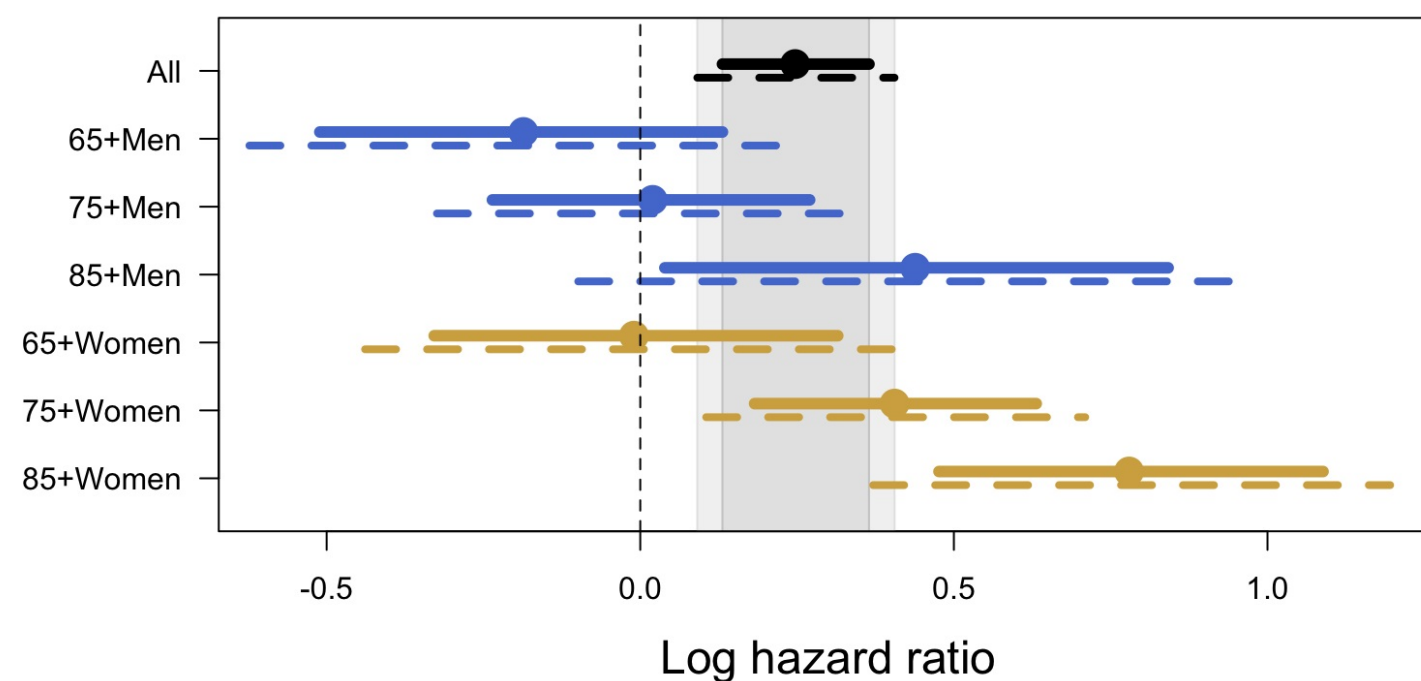
Correlation is a problem:

*“Lower-dose recipients were more likely to be older, to be receiving home healthcare or home oxygen, and to have higher CHADS<sub>2</sub> and HAS-BLED scores”*



# Subgroup trouble

*“Increased risk of major GI bleeding with dabigatran appeared to be restricted to women aged 75+ and to men aged 85+”*



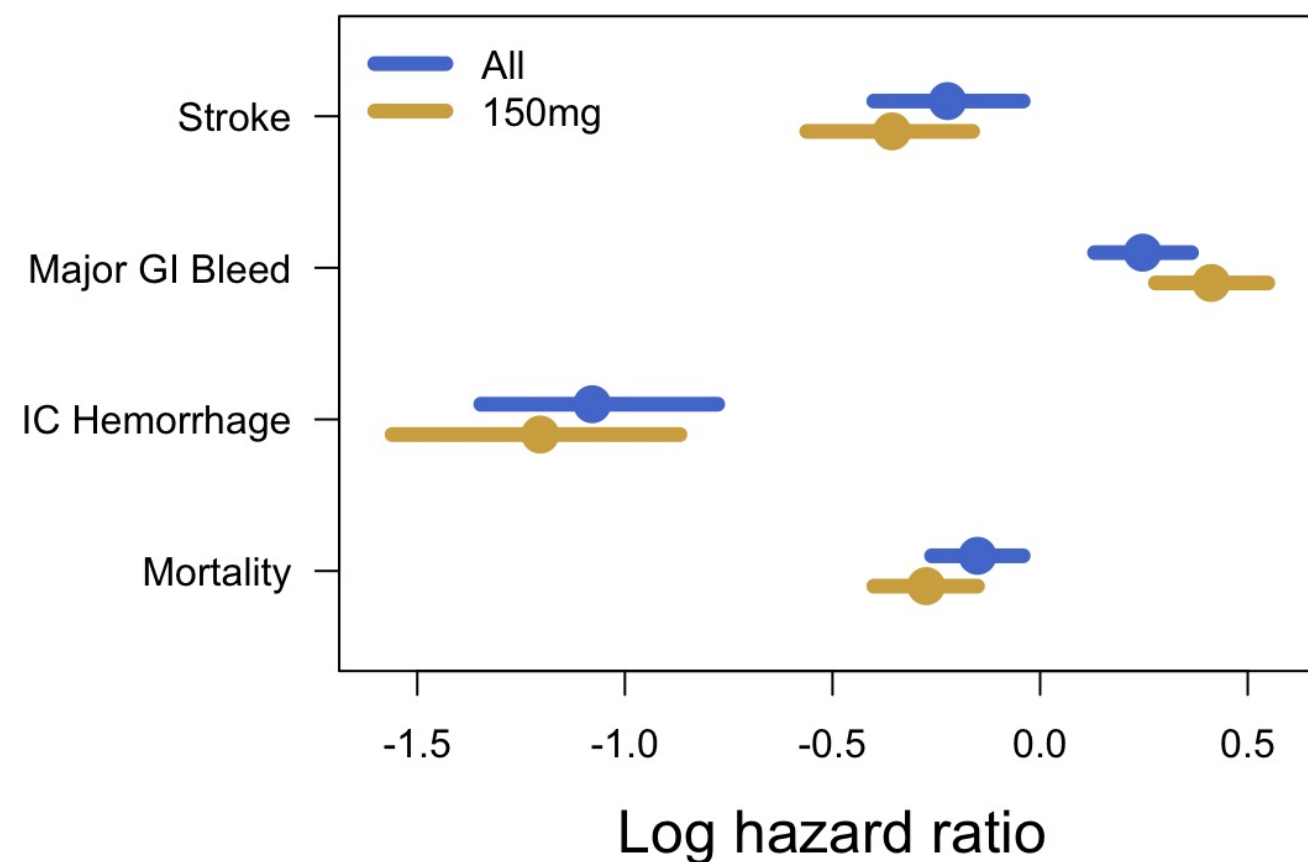
- CIs are not independent
- No adjustment for multiple testing
- Dashed lines assume only 6 tests

*“This shift in hazard ratio between younger and older women presented a **statistically significant** interaction”*



# Subgroup trouble

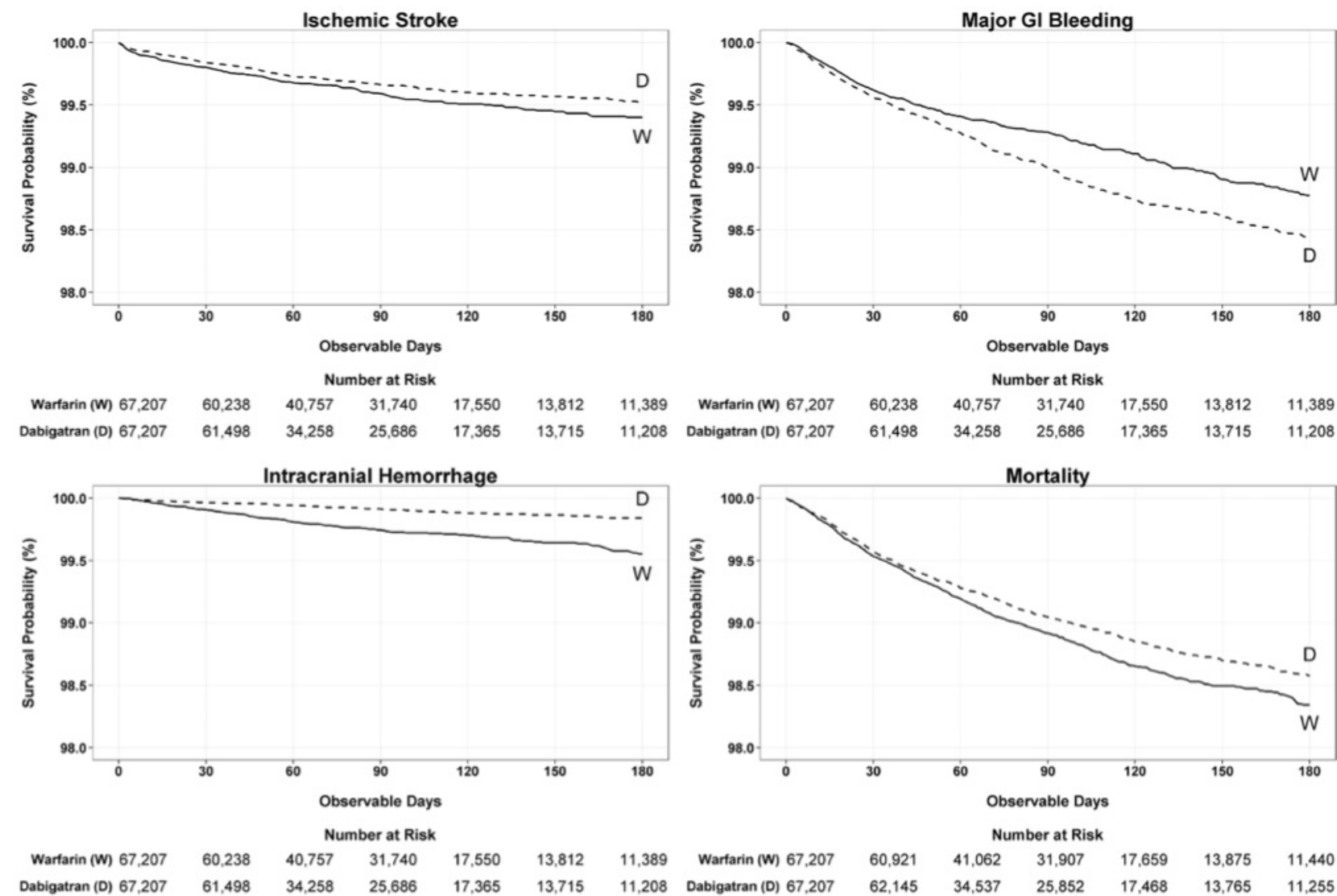
*“The magnitude of effect for each outcome was greater in the subgroup treated with dabigatran 150 mg twice daily compared with the main analysis”*



- Really?
- Where is the evidence?



# Kaplan-Meyer plots



**Figure.** Kaplan-Meier plots showing risk of ischemic stroke, major gastrointestinal (GI) bleeding, intracranial hemorrhage, and mortality in propensity score-matched cohorts treated with dabigatran (D; dotted line) or warfarin (W; solid line) for nonvalvular atrial fibrillation.

Where are the confidence intervals?



# Interpretation trouble

*“The absolute incidence of outcome events for both dabigatran and warfarin was greatest during the first 90 days of treatment, although the hazard ratios for these outcomes were constant over time.”*

1. Do we believe the effect is greater early in treatment?
2. Large population sizes lead more events
  - even when hazard is constant
3. Approximately 50% of population filled only one RX