Butterfly Effect in Studies Using Claims Data?

Some Small Perturbations in Study Design Lead to Differences in Causal Inference

Rima Izem, PhD
Office of Biostatistics
Center for Drug Evaluations and Research
Food and Drug Administration

Disclaimer

The views expressed are the presenter's and not necessarily those of the Food and Drug Administration

Workgroup Team

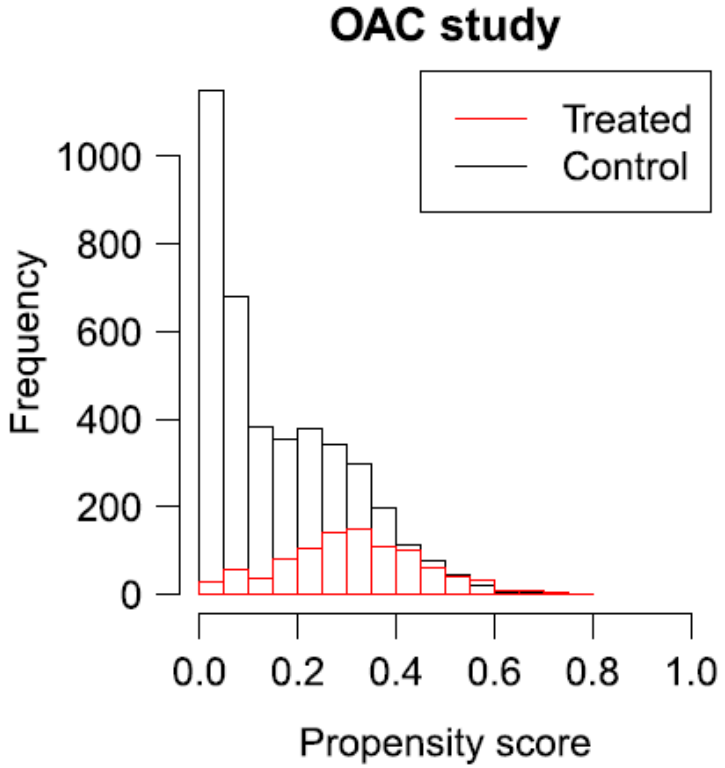
- FDA
  - Rima Izem, PhD
  - Michael Nguyen, MD

- Harvard Pilgrim Health Care Institute, Sentinel Operations Center
  - Judith C. Maro, PhD
  - Ting-Ying (Jane) Huang, PhD
  - Laura Shockro, BA
  - Laura Hou, MPH, MS
  - April Duddy, MS
  - Andrew Petrone, MPH
  - Ella Pestine, MPH

Outline

- Setting and Motivation
- Methods: Investigation in Two Phases
- Results
- Discussion

Setting: Drug Safety Studies

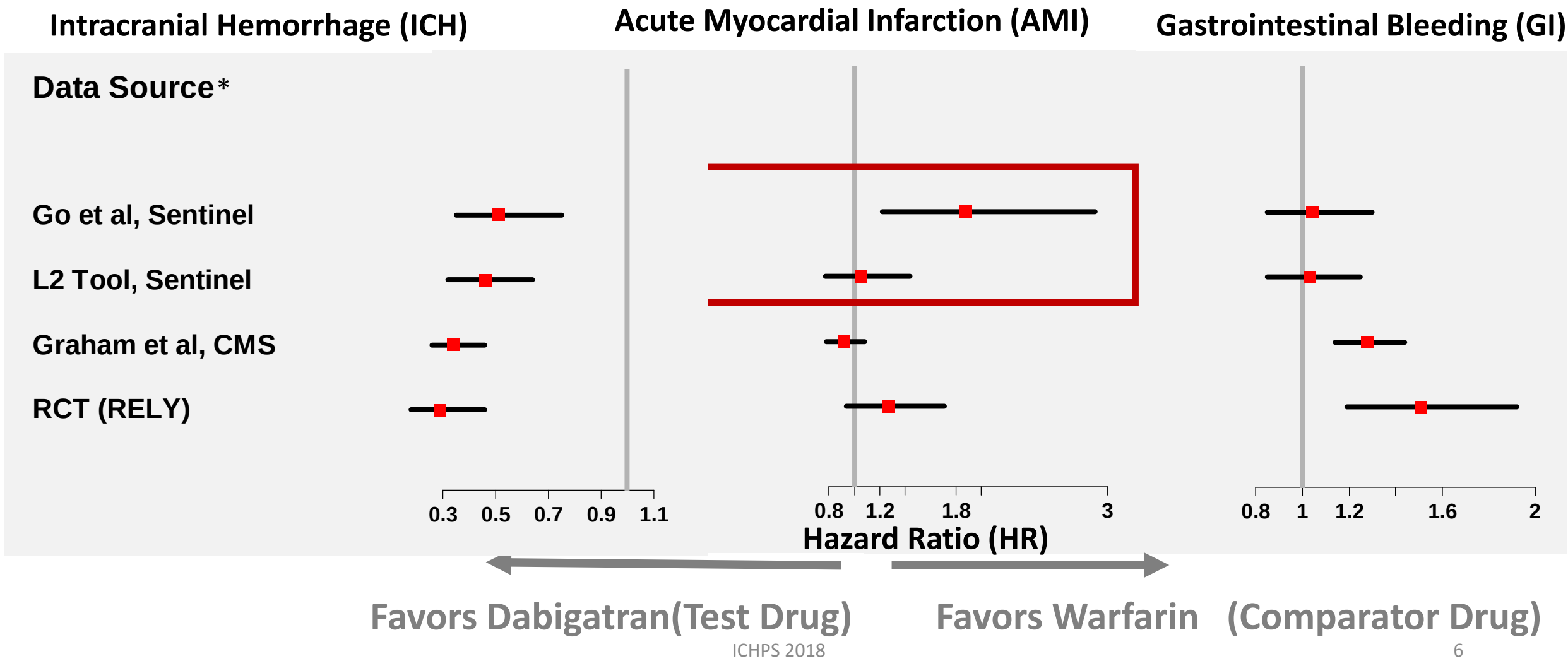


Source: Franklin et al (2017) – Statistics in Medicine

Design Elements: Claims Data, Retrospective New User Cohort, Chronic (Asymptomatic) Indication, 1:1 propensity score (PS) matching to control for confounding, short-term and long term outcomes (time to event analyses)

Motivation

Which Small Changes in Specifications Affect Risk Estimates in a Comparative Study?
How Do Small Changes in Specifications Impact Risk Estimates in a Comparative Study?



*Sources of estimates and spelled out acronyms are at the end of this presentation (slide 20)

Method: Identifying Factors Of Interest And Quantifying Their Impact

Phase I: Comparison of two *similar* codes which produced different risk estimates

Identified four *minor* specifications differences/factors impacting differences in cohort composition

A. **Day 0**
B. Inclusion/exclusion
C. **Stockpiling Algorithm**
D. Covariates in PS model
E. Stratification by matched set in Cox regression

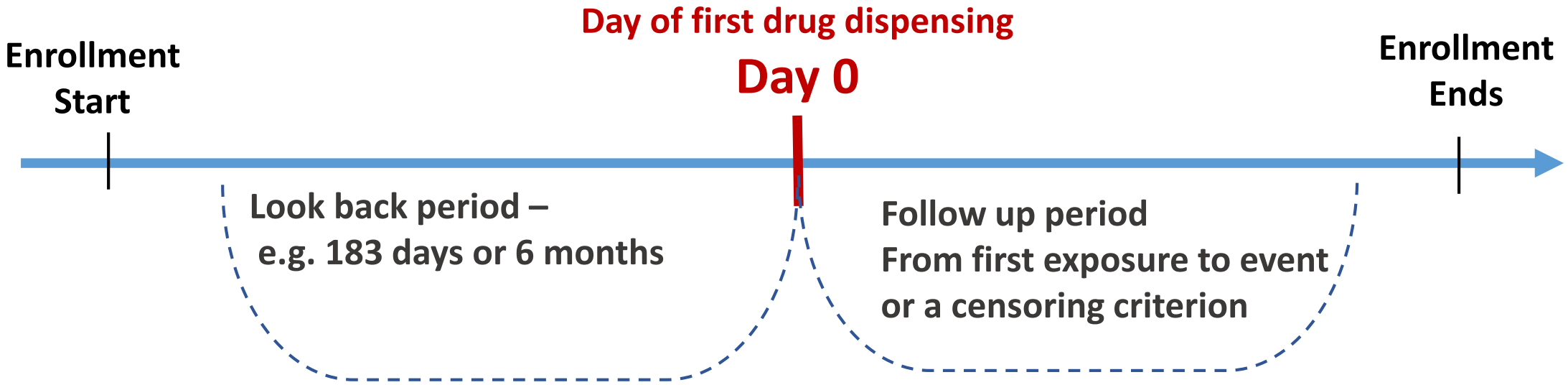
Phase II: Quantify impact of multiple factors on multiple outcomes on a test case

- Identify a test case in **MarketScan***
- Co-vary multiple factors with AMI
- Co-vary a select number of factors with GI bleed and ICH outcomes
- Quantify impact (summary level and **subject level**) from cohort composition to risk estimates

* Full reference: Truven Health MarketScan® Commercial Claims and Medicare Encounters Database

Factor A: Day 0

In Phase II of this investigation,
level A+: Day 0 in look back, not in follow up;
level A-: Day 0 in follow up, not in look back;

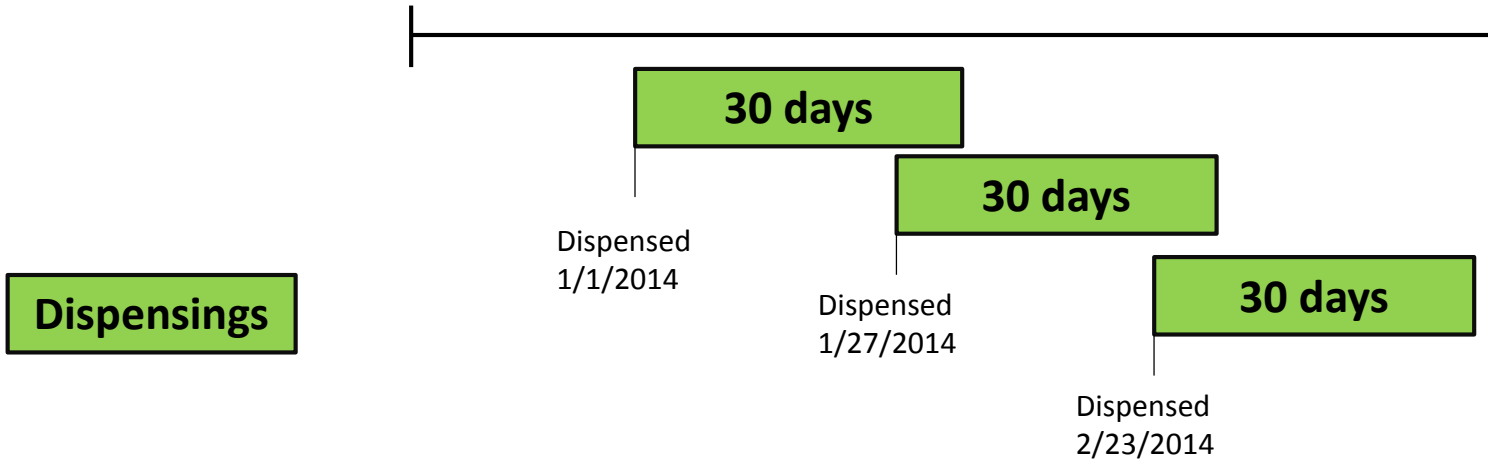


e.g. Go et al protocol "One or more diagnosis of atrial fibrillation or atrial flutter…any time before the first identified prescription for dabigatran or warfarin therapy during the study period"

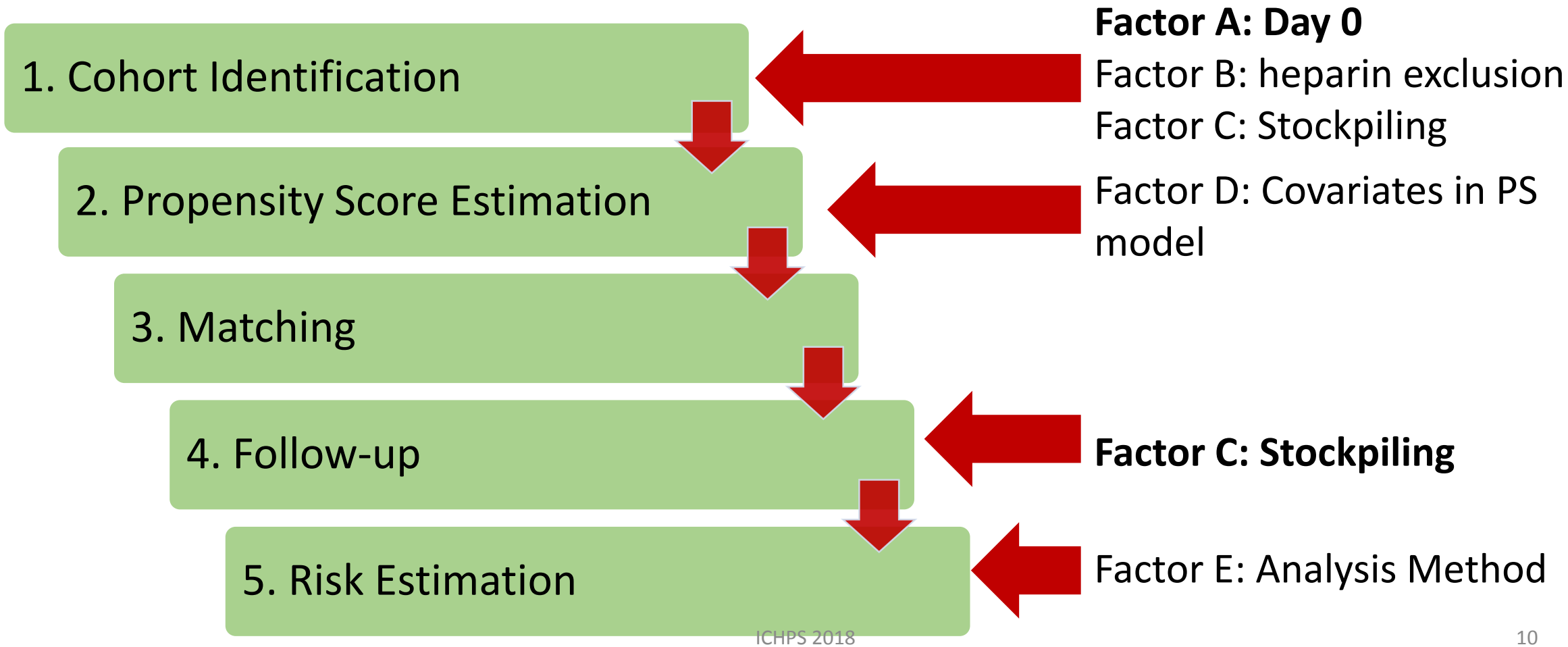
Is Day 0 in look back?

Factor C: Stockpiling, Vary Early Refill Specifications

In Phase II of this investigation,
level C+: same day Rx (sum) and consecutive Rx (add up all overlap);
level C-: same day Rx (max) and consecutive Rx (add up to 23% of overlap);

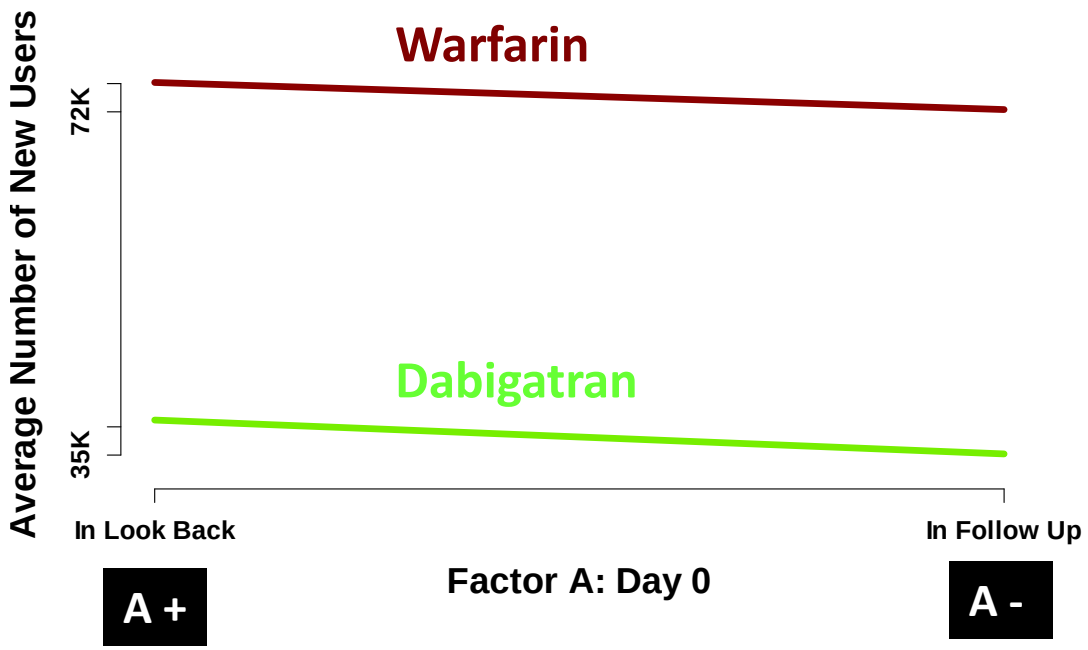


Expected Factor Impact on Risk Estimation Process

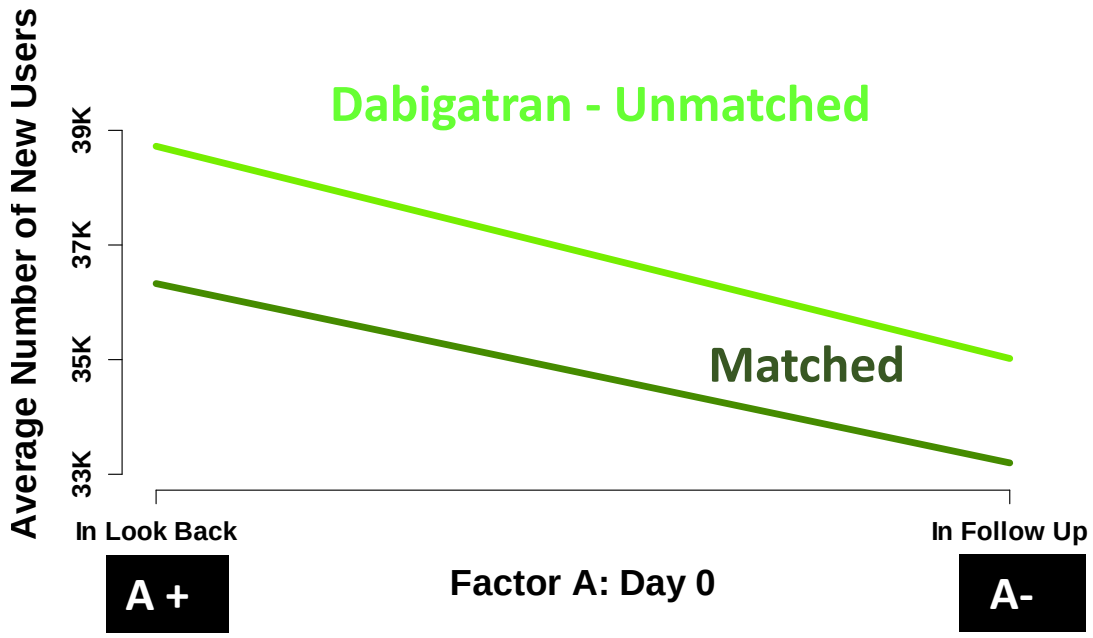


Results: Impact of Day 0 on Cohort Selection

Unmatched Cohort Size

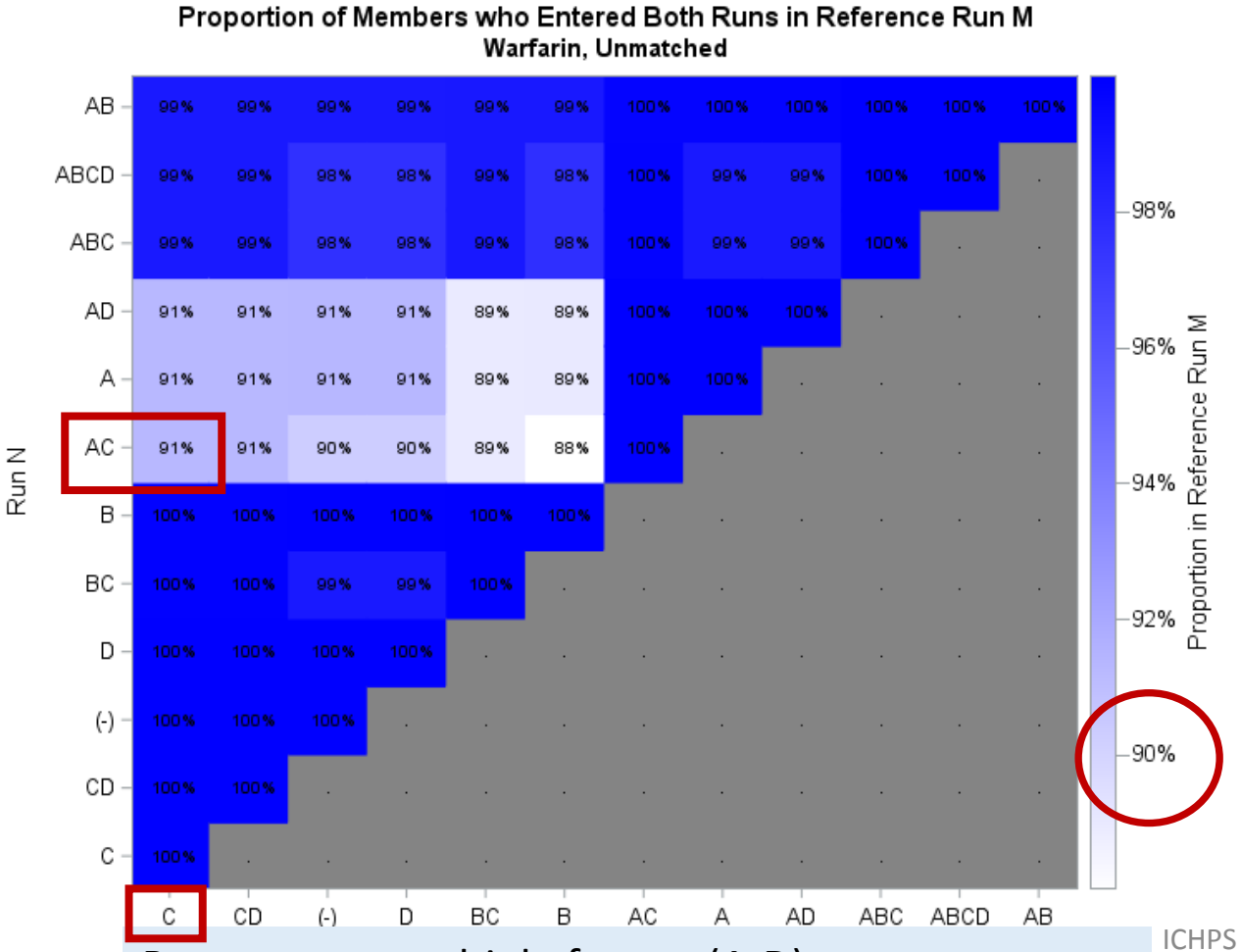


Matched Cohort Size

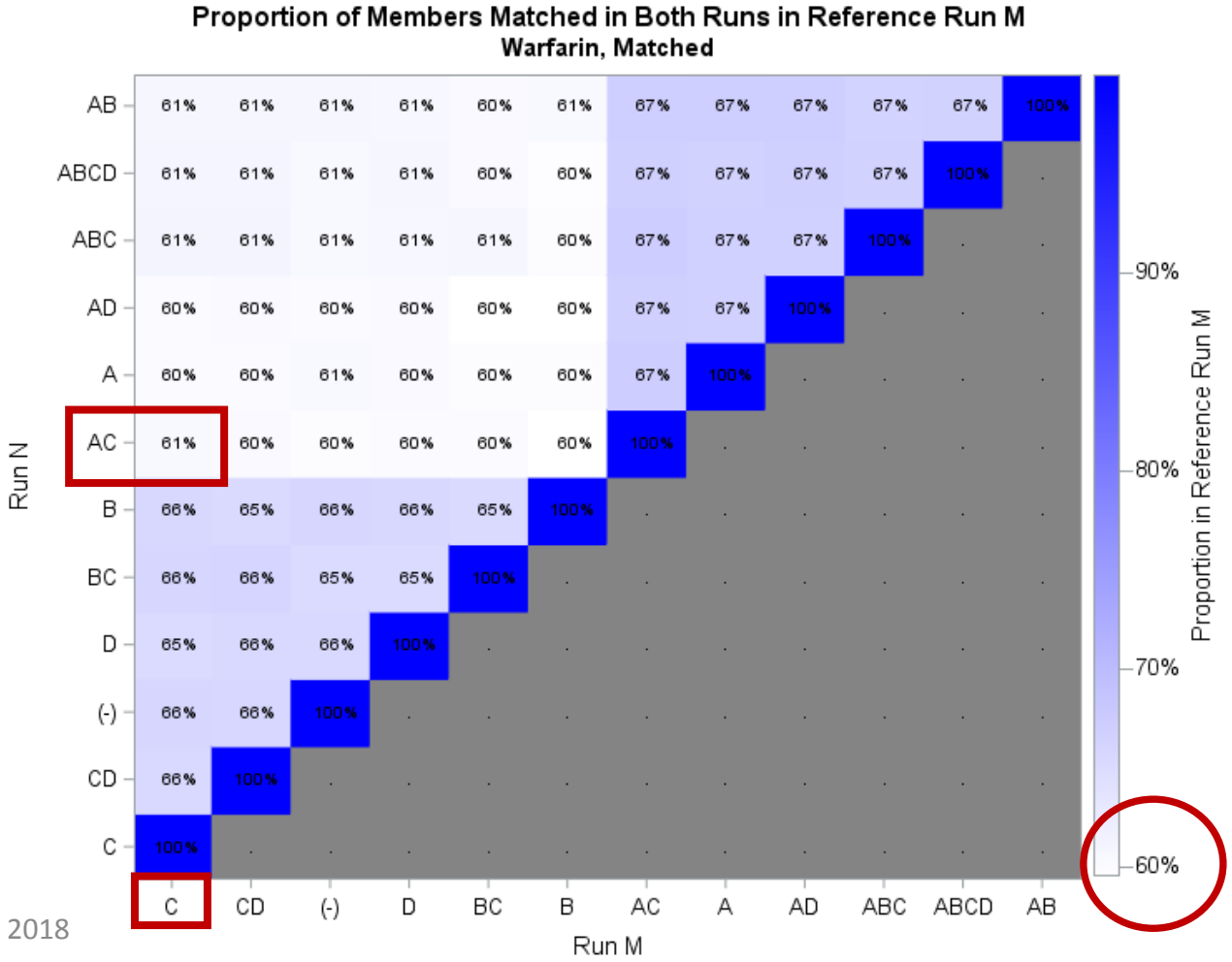


Results: Impact of Day 0 on Cohort Selection (continued)

Warfarin - Unmatched

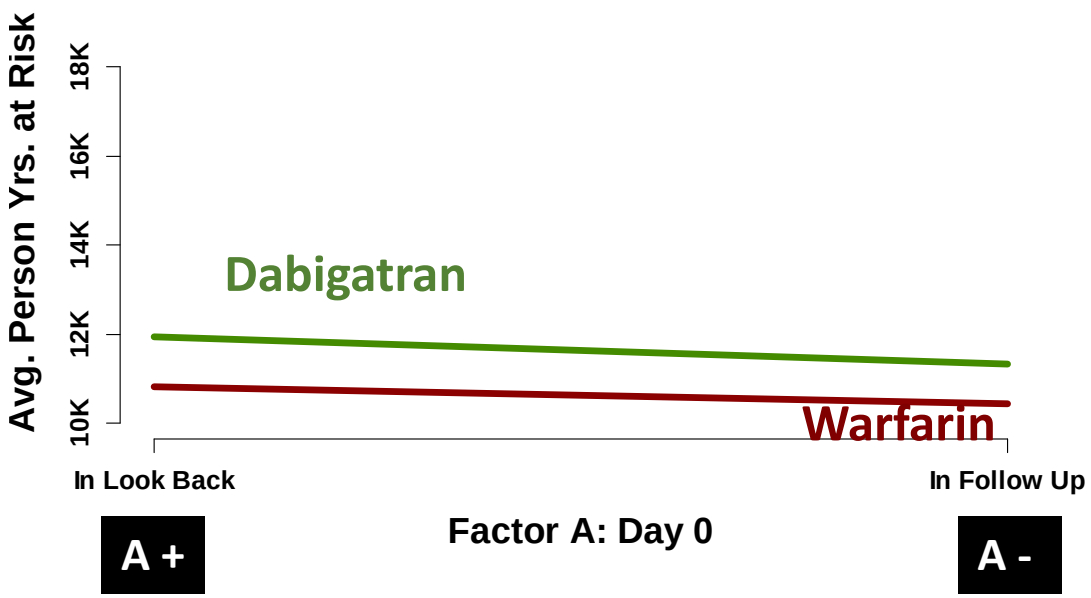


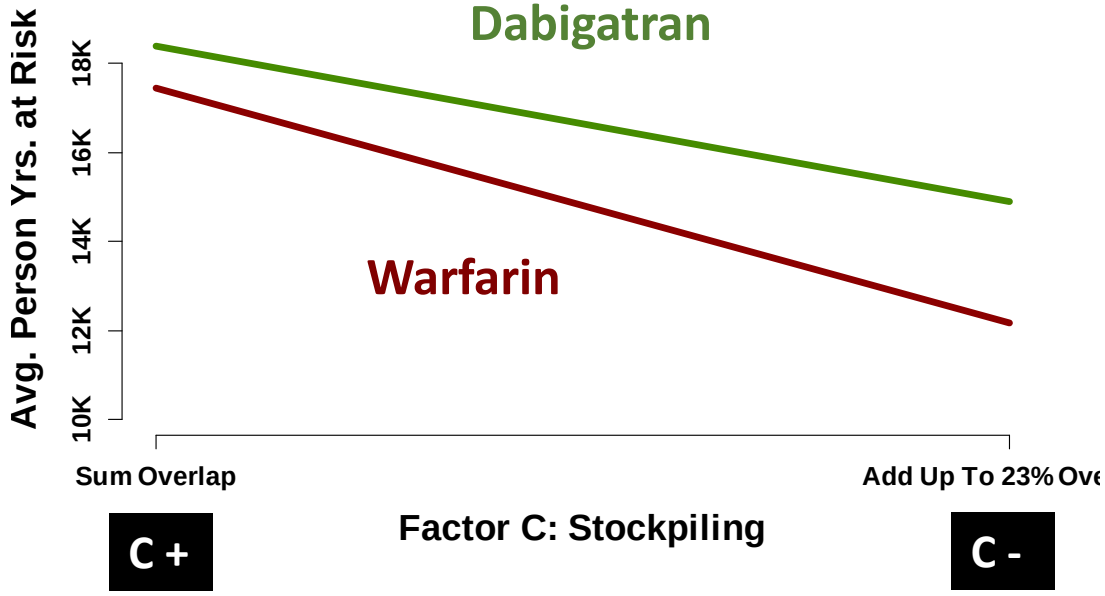
Warfarin - Matched



Runs co-vary multiple factors (A-D)

Results: Impact of Day 0 and Stockpiling on Patient-Years

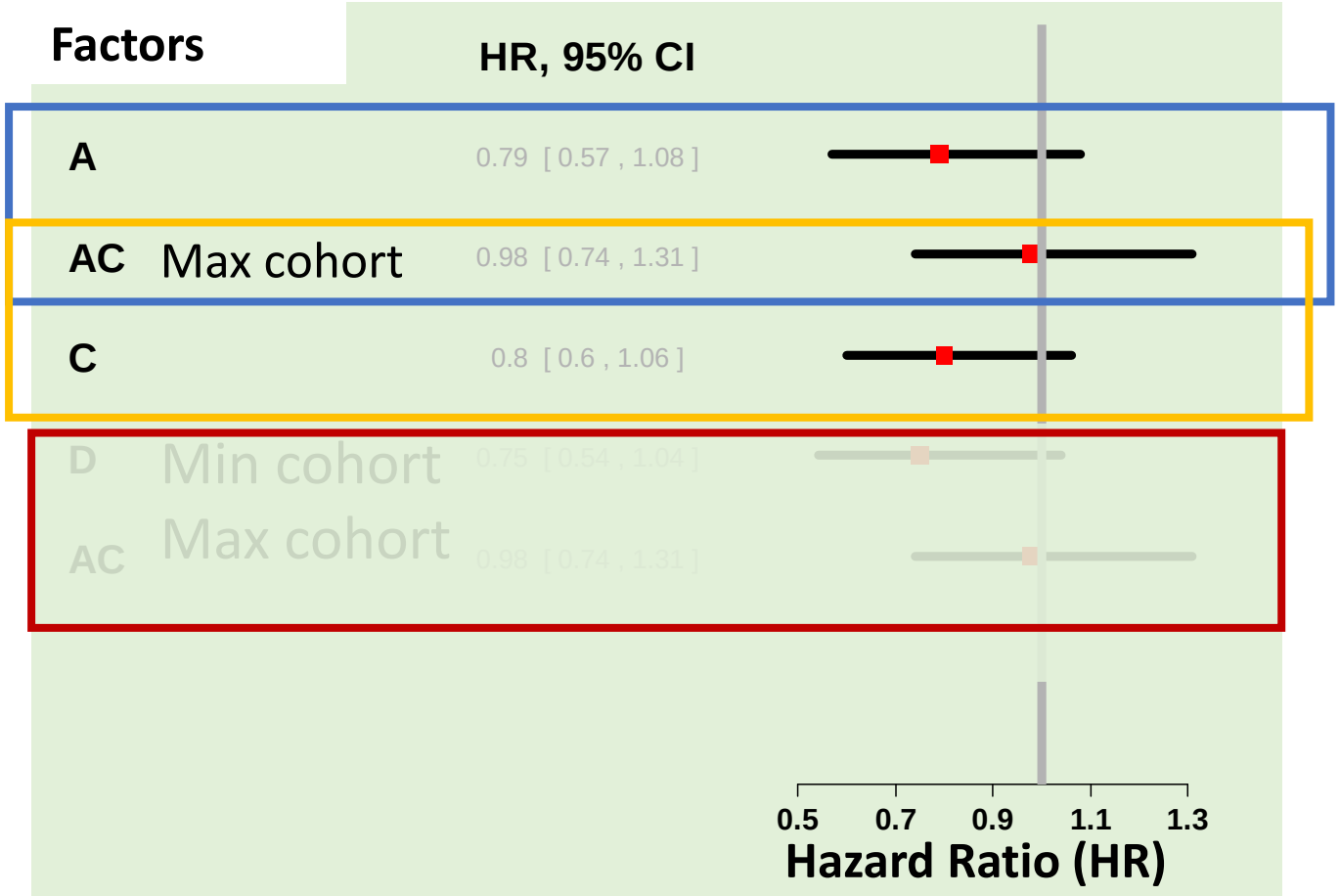




Results: Impact of Factors on Hazard Ratios for AMI

Recall Factors

A. Day 0
B. Heparin exclusion
C. **Stockpiling algorithm**
D. Covariates in PS model
E. Stratification by matched set in Cox regression

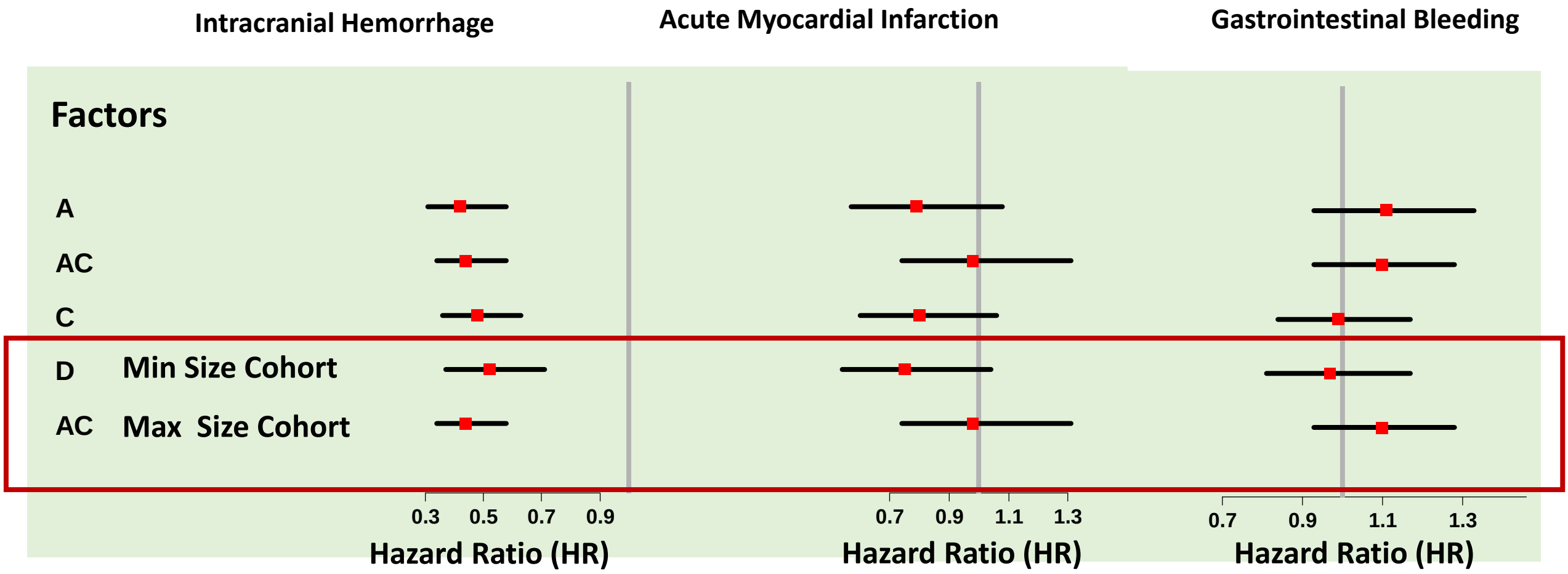


Change in stockpiling specification

Change in Day 0 specification

Change in multiple specifications for Day 0, stockpiling and covariates in PS model

Results: Impact of Day 0 and Stockpiling for Different Outcomes



Summary of Findings

- In safety study investigations, despite good study design practices led by team of experts, pre-specified protocols and statistical analysis plans and standardized programming, ***some*** unexpected differences were observed between thought to be similar analyses
- Our project identified small differences in interpretation of **Day 0** and **Stockpiling** specifications which could explain the differences
  - Including Day 0 in look back period resulted in a net increase of cohort size and person*time
  - Stockpiling specifications had a differential impact on person*time and incidence rates

Summary of Findings (Continued)

- Small changes to covariates in PS model did not greatly impact PS scores or matched cohort sizes
- Small changes in inclusion criteria (whether to exclude heparin use on index date) interacted with Day 0 choices before matching but had less impact after matching
- Even when small changes in factors did not impact overall matched sample size, they always impacted which warfarin subjects were matched

Limitations

- No two analyses in Truven Health MarketScan replicated the motivating difference in two results on AMI in Sentinel
- Changes of specifications were run on *one test case*, a pair of drugs and three outcomes and may not generalize to all other safety investigations
- Small changes in stockpiling impacted the titrated drug warfarin more than the fixed dose drug dabigatran—this differential impact may not generalize to comparison of two fixed dose drugs
- Small changes in specifications impacted risk estimates for *rare outcomes* but may have a smaller impact on more prevalent outcomes

Take Home

- **Replication**: In depth investigation would have been impossible with only “publication details” and needed access to source code. Shall we mandate publishing SAP, including statistical software codes to generate the cohort and analyze the data?
- **Specifications recommendations**: by default, include day 0 in look back in new user cohort studies. Explore stockpiling/include sensitivity analyses, especially for titrated drugs as the specifications may have differential impact
- **Standardization**: with more experience with safety studies in claims, specifications options and defaults will be standardized with reasons for defaults documented and downstream impact outlined

References for Background and Motivation

- Go et al study in Sentinel distributed database
  - Protocol available on https://www.sentinelinitiative.org/drugs/assessments/protocol-assessment-dabigatran-and-selected-safety-outcomes
  - Full reference: Go et al (2017). Outcomes of Dabigatran and Warfarin for Atrial Fibrillation in Contemporary Practice: the FDA Sentinel Program. *Annals of Internal Medicine doi:10.7326/M16-1157*.
- Level 2 (L2) tool in the Cohort Identification and Descriptive Analysis (CIDA) in Sentinel distributed database
  - Information on CIDA tools is available at https://www.sentinelinitiative.org/sentinel/surveillance-tools/routine-querying-tools/routine-querying-system), L2 controls for confounding
  - Sentinel reports are posted online at https://www.sentinelinitiative.org/drugs/assessments
- Graham et al study in Center for Medicaid and Medicare Services (CMS) database
  - Full reference: David Graham et al (2014) Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for non-valvular atrial fibrillation. *Circulation* October, 2014, doi: 10.1161/CIRCULATIONAHA.114.012.061.
- Randomized Evaluation of Long Term Anticoagulant Therapy (RELY) results
  - Results for GI bleeding and ICH are described in dabigatran drug label (Table 2, last updated in 2015)
  - Results for AMI are described in supplement of Connolly et al (2010) N Engl J Med 2010;363:1875-6.

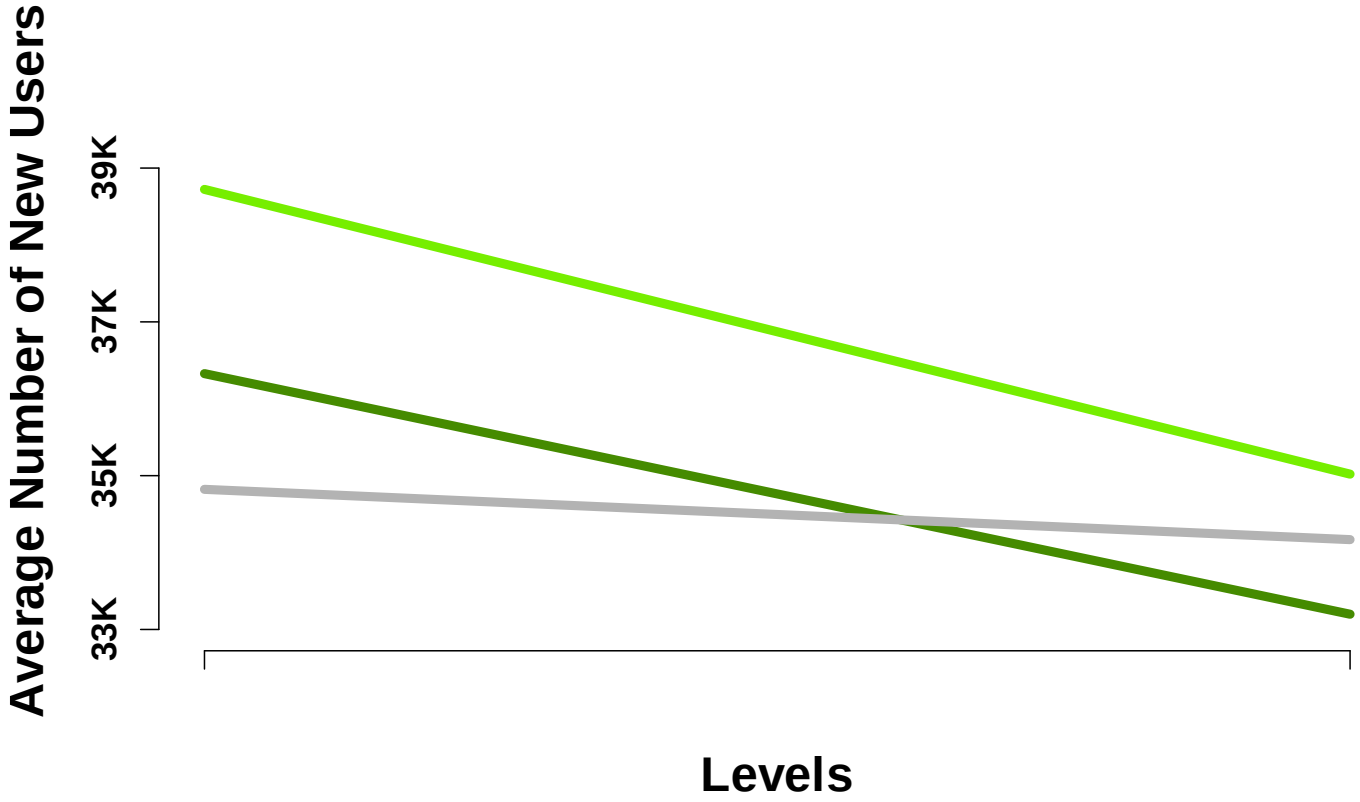
Thank you!

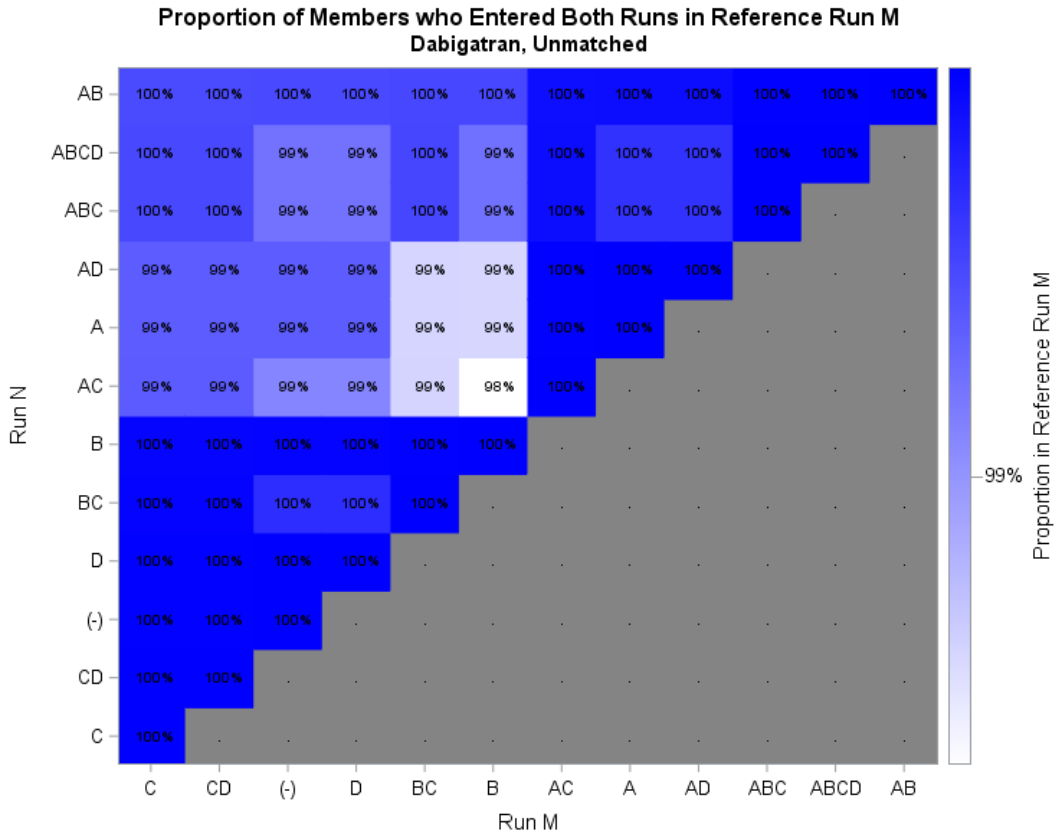
Any questions?

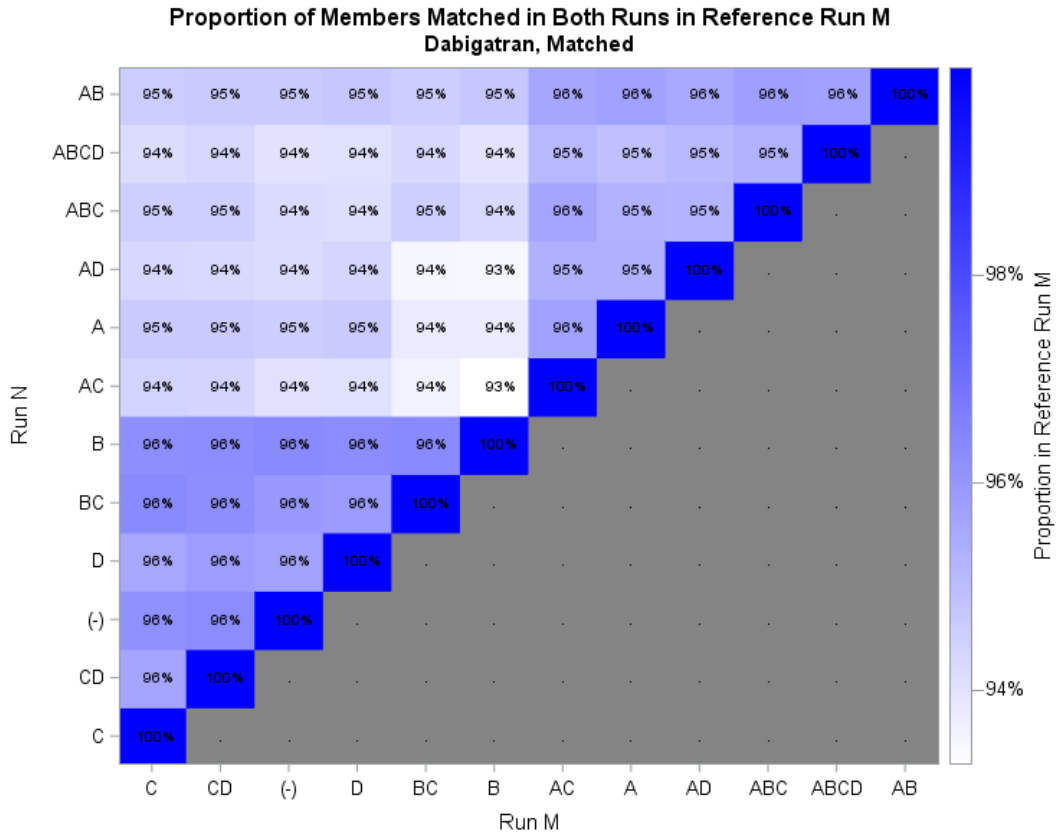
Rima.izem@fda.hhs.gov

Back up

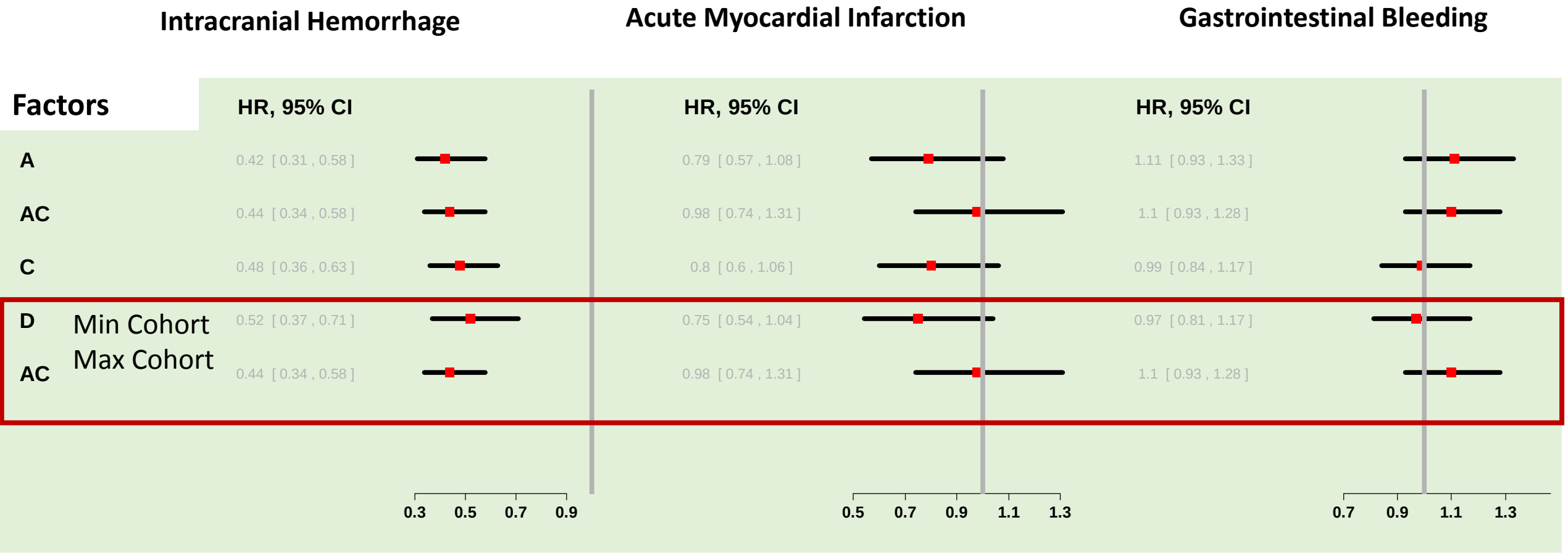
Factor A (dark green) vs. Factor D (grey) for matched cohorts




Proportion of Members who Entered Both Runs in Reference Run M
Dabigatran, Unmatched
Run N
AB 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100%
ABCD 100% 100% 99% 99% 100% 99% 100% 100% 100% 100% 100%
ABC 100% 100% 99% 99% 100% 99% 100% 100% 100% 100%
AD 99% 99% 99% 99% 99% 99% 100% 100% 100%
A 99% 99% 99% 99% 99% 99% 100% 100%
AC 99% 99% 99% 99% 99% 98% 100%
B 100% 100% 100% 100% 100% 100%
BC 100% 100% 100% 100% 100%
D 100% 100% 100% 100%
(-) 100% 100% 100%
CD 100% 100%
C 100%
C CD (-) D BC B AC A AD ABC ABCD AB
Run M
99%
Proportion in Reference Run M


Proportion of Members Matched in Both Runs in Reference Run M
Dabigatran, Matched
Run N
AB 95% 95% 95% 95% 95% 95% 96% 96% 96% 96% 96% 100%
ABCD 94% 94% 94% 94% 94% 94% 95% 95% 95% 95% 100%
ABC 95% 95% 94% 94% 95% 94% 96% 95% 95% 100%
AD 94% 94% 94% 94% 94% 93% 95% 95% 100%
A 95% 95% 95% 95% 94% 94% 96% 100%
AC 94% 94% 94% 94% 94% 93% 100%
B 96% 96% 96% 96% 96% 100%
BC 96% 96% 96% 96% 100%
D 96% 96% 96% 100%
(-) 96% 96% 100%
CD 96% 100%
C 100%
C CD (-) D BC B AC A AD ABC ABCD AB
Run M
98%
96%
94%
Proportion in Reference Run M

Results: Impact of Day 0 and Stockpiling for Different Outcomes



GI Bleeding – Age Effect

Appendix Table 3. Association of dabigatran versus warfarin use and outcomes in subgroup analyses.

	Hazard Ratio (95% Confidence Interval)						
	Age Group, years				Gender		
Outcome Type	**<65**	**65-74**	**75-84**	**≥85**	**Men**	**Women**	**Reduced Kidney Function**
	n=9438 pairs	n=7334 pairs	n=1287 pairs	n=2009 pairs	n=16,113 pairs	n=9143 pairs	n=1815 pairs
Ischemic stroke	1.09 (0.55-2.17)	1.10 (0.53-2.30)	0.87 (0.49-01.55)	1.00 (0.41-2.41)	0.86 (0.52-1.40)	1.00 (0.62-1.62)	0.27 (0.06-1.29)
Intracranial hemorrhage	0.39 (0.14-1.11)	**0.30 (0.12-0.74)**	0.68 (0.34-1.34)	0.67 (0.24-1.83)	**0.54 (0.32-0.94)**	**0.49 (0.24-0.99)**	0.72 (0.20-2.54)
Excluding trauma	0.53 (0.18-1.59)	**0.19 (0.05-0.65)**	0.58 (0.21-1.64)	0.65 (0.17-2.56)	0.51 (0.25-1.02)	**0.32 (0.13-0.83)**	–
Combined stroke	0.77 (0.44-1.37)	0.64 (0.37-1.12)	0.81 (0.52-1.26)	0.84 (0.43-1.62)	0.71 (0.49-1.03)	0.83 (0.56-1.23)	0.47 (0.18-1.21)
Excluding trauma	0.88 (0.49-1.58)	0.64 (0.35-1.15)	0.82 (0.50-1.37)	0.88 (0.42-1.84)	0.74 (0.49-1.11)	0.81 (0.53-1.23)	**0.20 (0.05-0.91)**
Major extracranial bleed	**0.51 (0.30-0.87)**	0.69 (0.46-1.04)	1.20 (0.86-1.68)	1.60 (0.96-2.69)	1.01 (0.76-1.34)	**0.73 (0.54-0.99)**	1.59 (0.93-2.72)
Gastrointestinal	0.59 (0.32-1.07)	0.81 (0.52-1.24)	**1.47 (1.01-2.14)**	**1.84 (1.05-3.20)**	1.26 (0.92-1.73)	0.78 (0.57-1.07)	**1.91 (1.04-3.51)**
Non-gastrointestinal	**0.11 (0.03-0.36)**	**0.12 (0.03-0.50)**	**0.29 (0.14-0.61)**	0.33 (0.07-1.63)	**0.20 (0.10-0.39)**	**0.22 (0.08-0.58)**	0.52 (0.17-1.56)
Myocardial infarction	2.13 (0.98-4.66)	0.97 (0.06-15.56)	**4.09 (1.39-12.03)**	**5.25 (1.17-23.60)**	**2.06 (1.17-3.64)**	1.69 (0.84-3.38)	2.18 (0.20-24.18)

Source: Go et al (2017)

GI Bleed – Age Effect (continued)

Table 3. The effect of age and gender on the risk of ischemic stroke, intracranial hemorrhage, major gastrointestinal bleeding and mortality in propensity score matched cohorts treated with dabigatran or warfarin for non-valvular atrial fibrillation. Warfarin was the reference group.*

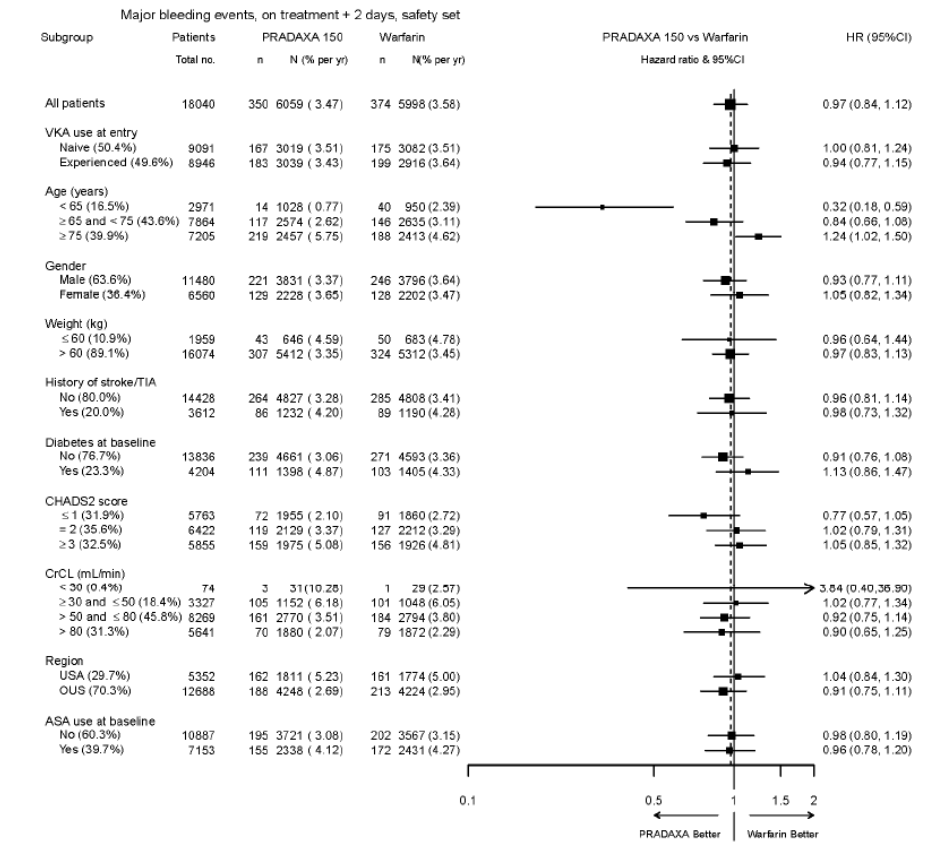
	Age-group (n)	Men Hazard ratio (95% CI)	Women Hazard ratio (95% CI)
Ischemic stroke			
	65-74 (55,761)	0.69 (0.42-1.14)	0.81 (0.51-1.31)
	75-84 (57,345)	0.98 (0.64-1.51)	0.89 (0.64-1.26)
	≥ 85 (21,308)	0.89 (0.41-1.90)	0.60 (0.40-0.91)
Intracranial hemorrhage			
	65-74 (55,761)	0.32 (0.15-0.68)	0.13 (0.04-0.44)
	75-84 (57,345)	0.27 (0.14-0.50)	0.59 (0.35-0.98)
	≥ 85 (21,308)	0.51 (0.18-1.48)	0.26 (0.12-0.56)
Major GI bleeding			
	65-74 (55,761)	0.83 (0.60-1.14)	0.99 (0.72-1.37)
	75-84 (57,345)	1.02 (0.79-1.31)	1.50 (1.20-1.88)
	≥ 85 (21,308)	1.55 (1.04-2.32)	2.18 (1.61-2.97)
Mortality			
	65-74 (55,761)	0.81 (0.62-1.05)	0.72 (0.52-0.99)
	75-84 (57,345)	0.73 (0.58-0.92)	0.82 (0.65-1.03)
	≥ 85 (21,308)	0.92 (0.64-1.33)	1.24 (0.96-1.60)

* Age-gender specific incidence rates of outcome events for the dabigatran and warfarin cohorts are shown in Supplemental Tables 4 and 5.

Source: Graham et al (2015)

GI Bleeding – Age Effect

Figure 1 Adjudicated Major Bleeding by Baseline Characteristics Including Hemorrhagic Stroke Treated Patients



Source: dabigatran (2015) label

Impact of Changes in Specifications

Specifications Examples

Length of look back period
Inclusion/Exclusion

1. Cohort Identification

Propensity score model
Covariates

2. Propensity Score Estimation

Caliper
Matching ratio

3. Matching

Stockpiling algorithm
Censoring criteria

4. Follow-up

Outcome model

5. Risk Estimation