Using Time-Dependent Propensity Scores to Correct for Differential Depletion of Susceptibles when Estimating Hazard Ratios in Time-to-Event Data

Richard Wyss PhD, MS,

ICPE 2018

August 24, 2018

Disclosures

- This work was funded by Sentinel Contract number: HHSF223200910006I.

- No other conflicts of interest

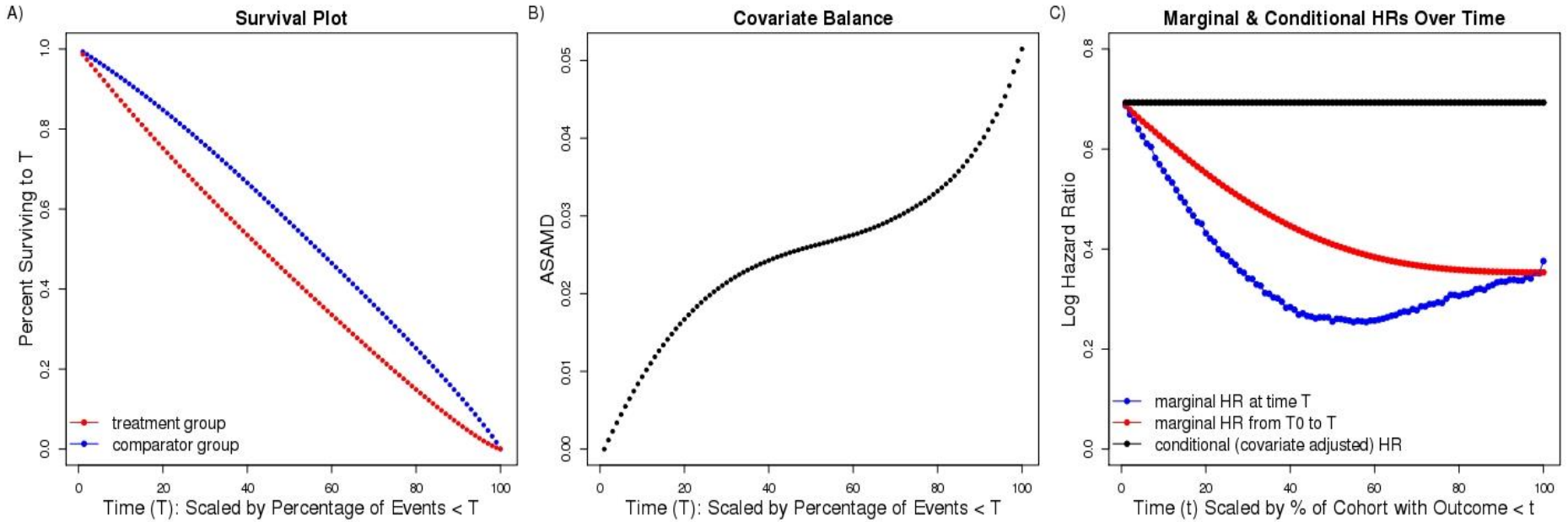
Background & Objectives

Background

- Different types of treatment effects:
  - Conditional treatment effect (patient specific)
  - Marginal or population averaged treatment effects (ATT, ATE).
- With no treatment effect heterogeneity:
  - Marginal treatment effects are equivalent to the conditional treatment effect for many measures of effect (e.g., risk ratio, rate ratio, risk difference).
  - However, marginal HR can differ from the conditional HR as a result of differential depletion of susceptibles.

Introduction

- Differential depletion of susceptibles
  - Occurs when the treatment affects outcome risk.
  - Results in an imbalance in baseline covariates across treatment groups over time.
- In these settings, there is a built in selection bias when estimating the marginal hazard ratio (Hernan 2010).
- The marginal (population averaged) HR will change over the course of follow-up and diverge from the conditional (covariate adjusted) hazard ratio.



- **Figure 1a** shows the survival curves of a treated and comparator group in a simulated population where the treatment of interest increases outcome risk, resulting in the treated population experiencing the outcome event at a faster rate relative to the comparator group.
- **Figure 1b:** As high-risk individuals are differentially depleted, the population at risk becomes unbalanced on baseline covariates over time with an increase in the average standardized absolute mean difference (ASAMD) in covariates across treatment groups.
- **Figure 1c:** As the population at risk becomes unbalanced on baseline covariates over time, marginal (population averaged) and conditional (covariate adjusted) hazard ratios diverge from each other.

Background

- In the presence of differential depletion of susceptibles, the marginal hazard ratio systematically moves away from the conditional hazard ratio towards the null in a way that is prone to mislead intuitions about the magnitude of the treatment effect.
- Therefore, if differential depletion of susceptibles is substantial, valid estimation of conditional hazard ratios may provide valuable insight for evaluating drug safety.
  - Estimating conditional hazard ratios through direct covariate adjustment, however, is not always possible as the fitting of high-dimensional outcome models can be problematic within distributed data environments.

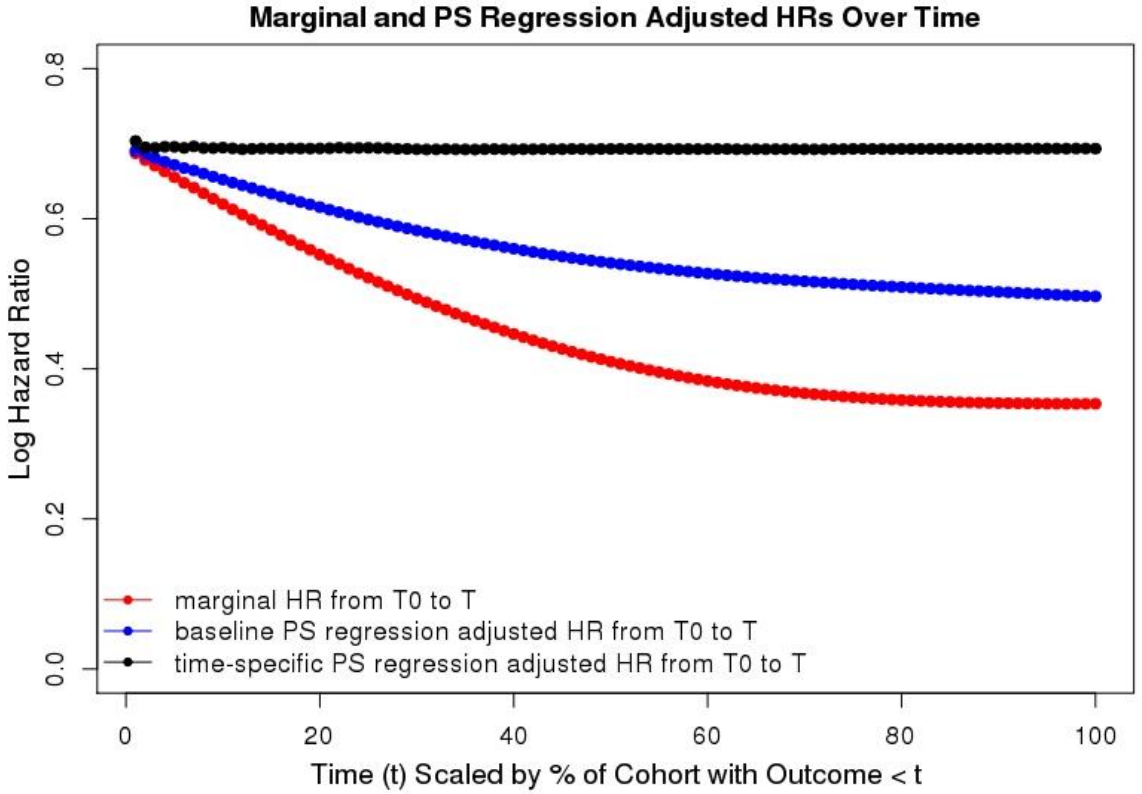
Objectives

- In this study, we propose a simple method that conditions on time-dependent propensity scores to correct for covariate imbalances caused by differential depletion of susceptibles.
- In the absence of effect heterogeneity and assuming correct model specification, the proposed method estimates the same conditional hazard ratio as would be obtained by adjustment for the covariates individually.
- We then use “plasmode” simulations to compare the performance of various methods for targeting marginal and conditional HRs to provide insight on the impact of differential depletion of susceptibles in more realistic settings.

Methods

Methods

- Time-Dependent Propensity Scores
  - For propensity score-based methods to consistently estimate conditional hazard ratios, the propensity score needs to vary over time as the population at risk changes due to depletion of susceptibles.
  - This is done by simply re-estimating the PS within the population that is still at risk at different time points over the course of follow-up.
    - Each of the time-dependent PS's is a function of baseline covariates. Keep in mind that we are not updating the values of the baseline covariates – only the function of them that predicts treatment status, conditional on still being at risk.
    - We then condition on the time-dependent PS's as a time-dependent covariate in the Cox outcome model.



- A simulated example of treatment arms that are balanced when t=0, then individuals across treatment groups are depleted differentially over time due to the treatment effect.
- Estimated hazard ratios that are produced from conditioning on the baseline propensity score fall somewhere between the marginal and conditional HR.
- In contrast, estimated hazard ratios that are produced from conditioning on a time-dependent propensity score are unbiased for the conditional hazard ratio

Methods

- "Plasmode" simulation
  - *Empirical data*: initiators of dabigatran versus warfarin in the Truven Marketscan Database between October 2010 (the month of dabigatran's approval in the US) and December 2013.
    - 79,265 individuals with 69 baseline covariates. When dichotomizing multi-level categorical variables, these baseline covariates account for 92 binary terms
  - Of the 92 binary terms, we identified the top 50% (46 terms) with the strongest confounding effect as measured by the Bross formula.
  - Multivariate associations between these selected variables with the outcome were simulated to be representative of those observed in the study cohort.
  - Simulated datasets were then created by sampling, with replacement, 40,000 individuals from this dataset.

Methods

- In previous work, Gruber et al (2018) identified five parameters that impact the magnitude of differential depletion of susceptibles:
  1. Strength of the treatment effect.
  2. Proportion of individuals in the population who experience the outcome event.
  3. Correlation between the propensity score and disease risk score.
  4. Amount of censoring (even if uninformative).
  5. Strength of covariate effects on the outcome.
- Here, we considered a range of scenarios where we varied four of these parameters (parameters 1 through 4 listed above).
  - We want to assess the practical impact of each of these parameters on depletion of susceptibles and the estimation of marginal and conditional HR's in settings reflective of real world data.

Sentinel

Table 1. Description of Simulation Scenarios				
Scenario[a]	Conditional HR	Log Conditional HR	Correlation btw PS and DRS[b]	Outcome Incidence
1	2	0.693	Moderate	50%
2	2	0.693	Moderate	20%
3	2	0.693	Moderate	10%
4	2	0.693	Moderate	5%
5	2	0.693	Moderate	2%
6	3	1.099	Moderate	50%
7	3	1.099	Moderate	20%
8	3	1.099	Moderate	10%
9	3	1.099	Moderate	5%
10	3	1.099	Moderate	2%
11	2	0.693	Weak	50%
12	2	0.693	Weak	20%
13	2	0.693	Weak	10%
14	2	0.693	Weak	5%
15	2	0.693	Weak	2%
16	3	1.099	Weak	50%
17	3	1.099	Weak	20%
18	3	1.099	Weak	10%
19	3	1.099	Weak	5%
20	3	1.099	Weak	2%

Methods

- **Methods for confounding adjustment:**
  1. 1-to-1 propensity score matching at baseline.
  2. 1-to-1 propensity score matching at baseline with stratified analyses.
  3. Adjusting for the baseline propensity score directly within a Cox regression outcome model.
  4. Adjusting for time-specific propensity scores directly within a Cox regression outcome model.
- For comparison, we also estimated the unadjusted (crude) hazard ratio and the estimated hazard ratio obtained from fitting the true covariate adjusted outcome model.

Selected Results

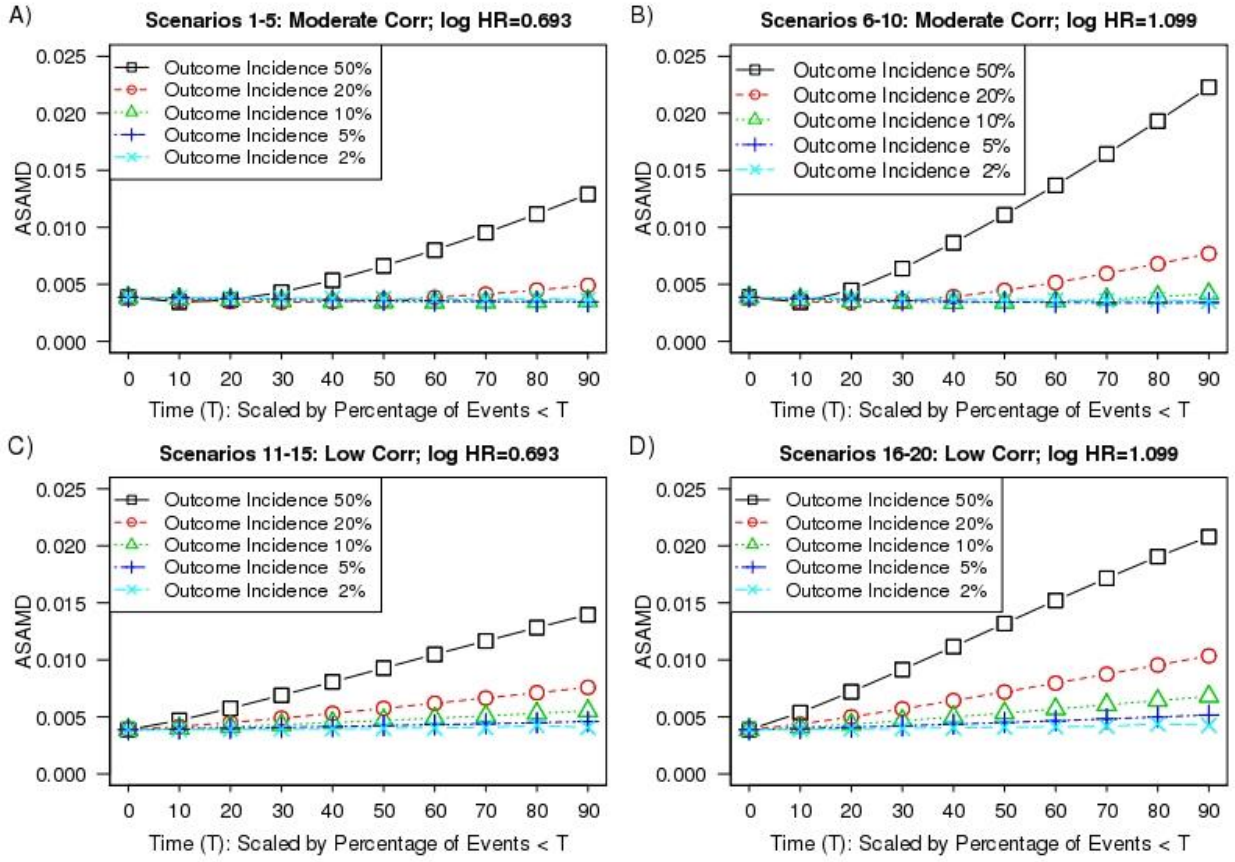


Figure 3. Covariate balance across treatment groups (measured by the average standardized absolute mean difference (ASAMD) in baseline covariates).

- Figures show that the amount of imbalance caused by differential depletion of susceptibles increases with higher outcome incidence, and increases with the strength of the treatment effect. When outcome incidence is low, the impact of differential depletion of susceptibles on covariate balance is small.

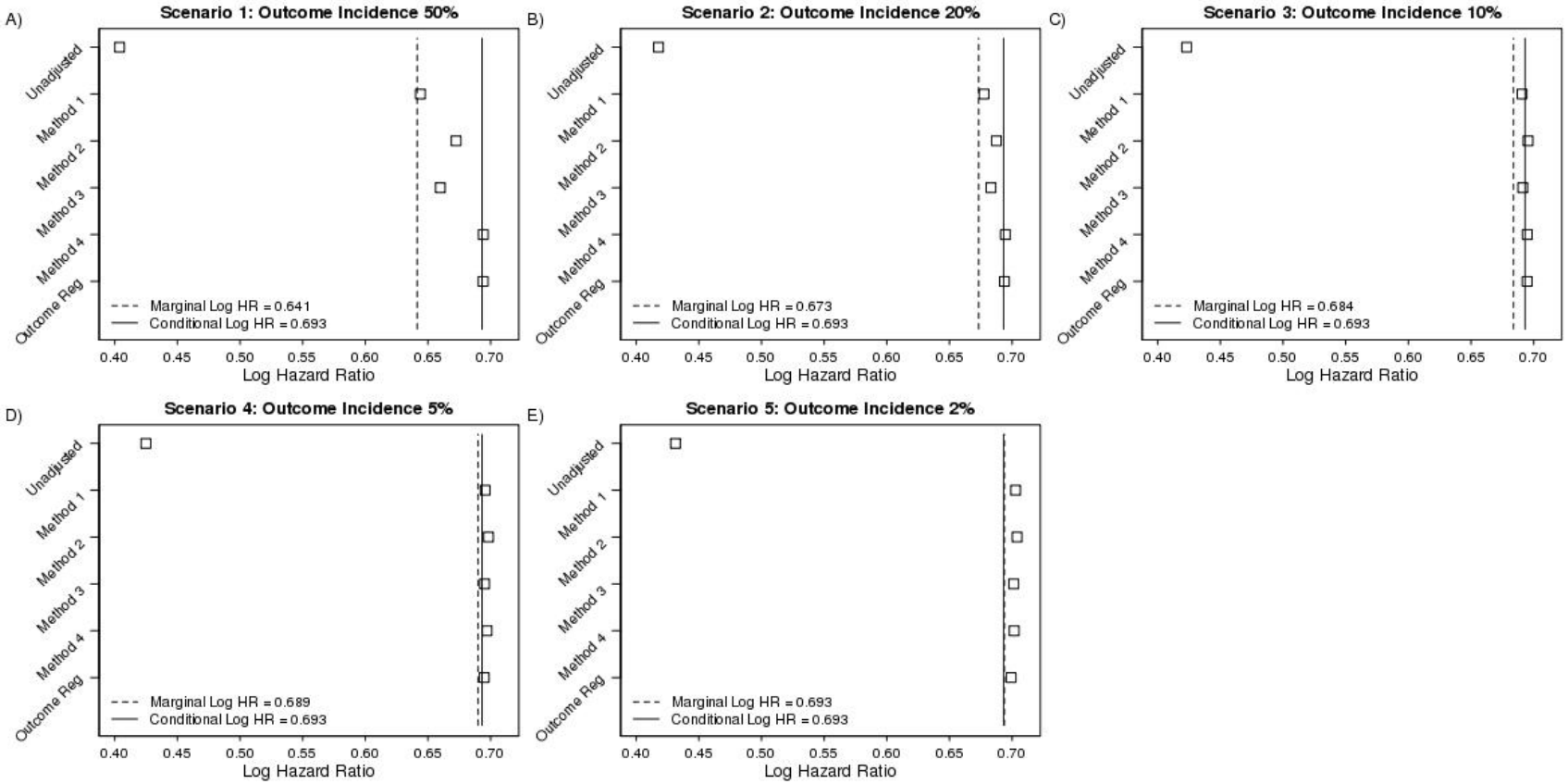
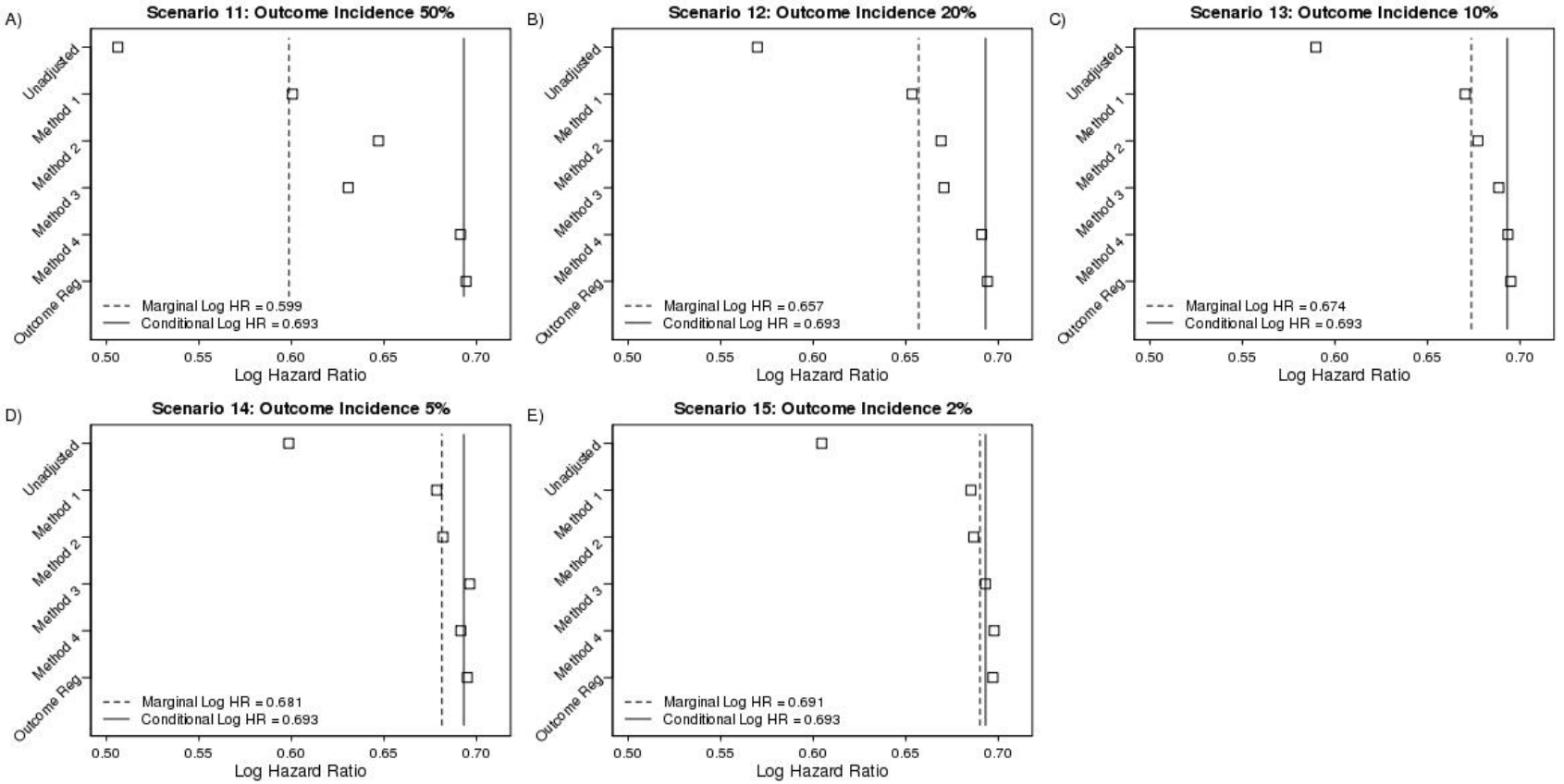


Figure 4 (moderate correlation btw PS & DRS). Estimated log hazard ratios for Scenarios 1 through 5.

- For scenarios with high outcome incidence, propensity score methods that condition on the baseline propensity score (methods 2 and 3) produce biased estimates of the conditional HR.
- Adjusting for a time-dependent PS (Method 4) produces approximately unbiased estimates of the conditional HR.
- When outcome incidence is low, the impact of differential depletion of susceptibles is small and the conditional and marginal HR are approximately equal.



was equal to 3 and the correlation between the propensity score and disease risk score was moderate at a val

Figure 5 (low correlation btw PS & DRS). Estimated log hazard ratios for Scenarios 11 through 15.

- When compared with Figure 4, the difference between the marginal and conditional hazard ratio is greater due to the low correlation between the PS and DRS.
- However, overall patterns are similar to Figure 4. In particular, when outcome incidence is low, the impact of differential depletion of susceptibles is small and the conditional and marginal HR are approximately equal.

Discussion, Limitations, & Conclusions

Discussion

- For the scenarios evaluated in this study, we found that the impact of differential depletion of susceptibles was negligible when the outcome incidence was low (<10% when correlation between the propensity and risk score was moderate and <5% when the correlation between the propensity and risk score was weak).
- When outcome incidence was moderate to high, however, the impact of differential depletion of susceptibles could be substantial with large differences between the marginal and conditional hazard ratios.
- Adjusting for time-dependent propensity scores as a time-dependent covariate directly within the outcome Cox model successfully adjusted for covariate imbalances over time providing unbiased estimates of conditional hazard ratios.

Limitations

- In this study, we did not consider settings involving treatment effect heterogeneity. In the presence of heterogeneity, the conditional HR is no longer constant.
  - However, if heterogeneity in the treatment effect is moderate, it can be useful to estimate the conditional hazard ratio “as if” it were constant to provide insight on the overall magnitude of the treatment effect and safety of the medical product.
- In this study, we also did not consider time-varying confounding or informative censoring. In settings complicated by time-varying confounding and informative censoring the observed results may not generalize.
- Finally, alternative measures of effect that are not adversely impacted by differential depletion of susceptibles could be used for evaluating drug safety.

Conclusions

- Adjusting for time-specific propensity scores can correct for covariate imbalances over time that are caused by differential depletion of susceptibles.
- Conditioning on time-dependent propensity scores can provide an alternative to outcome regression modeling for estimating conditional hazard ratios in distributed data settings where fitting high-dimensional outcome models can be difficult
- However, when outcome events are rare—situations that are common in post-market safety surveillance—our results suggest that the impact of differential drop of out susceptibles will be small.

Acknowledgements

Brigham & Women's Hospital

- Richard Wyss (Lead)
- Rishi J. Desai
- Jessica M. Franklin
- Joshua J. Gagne
- Sebastian Schneeweiss
- Shirley V. Wang

Food & Drug Administration (FDA)

- Jacqueline M. Major
- Yueqin Zhao
- Esther H. Zhou

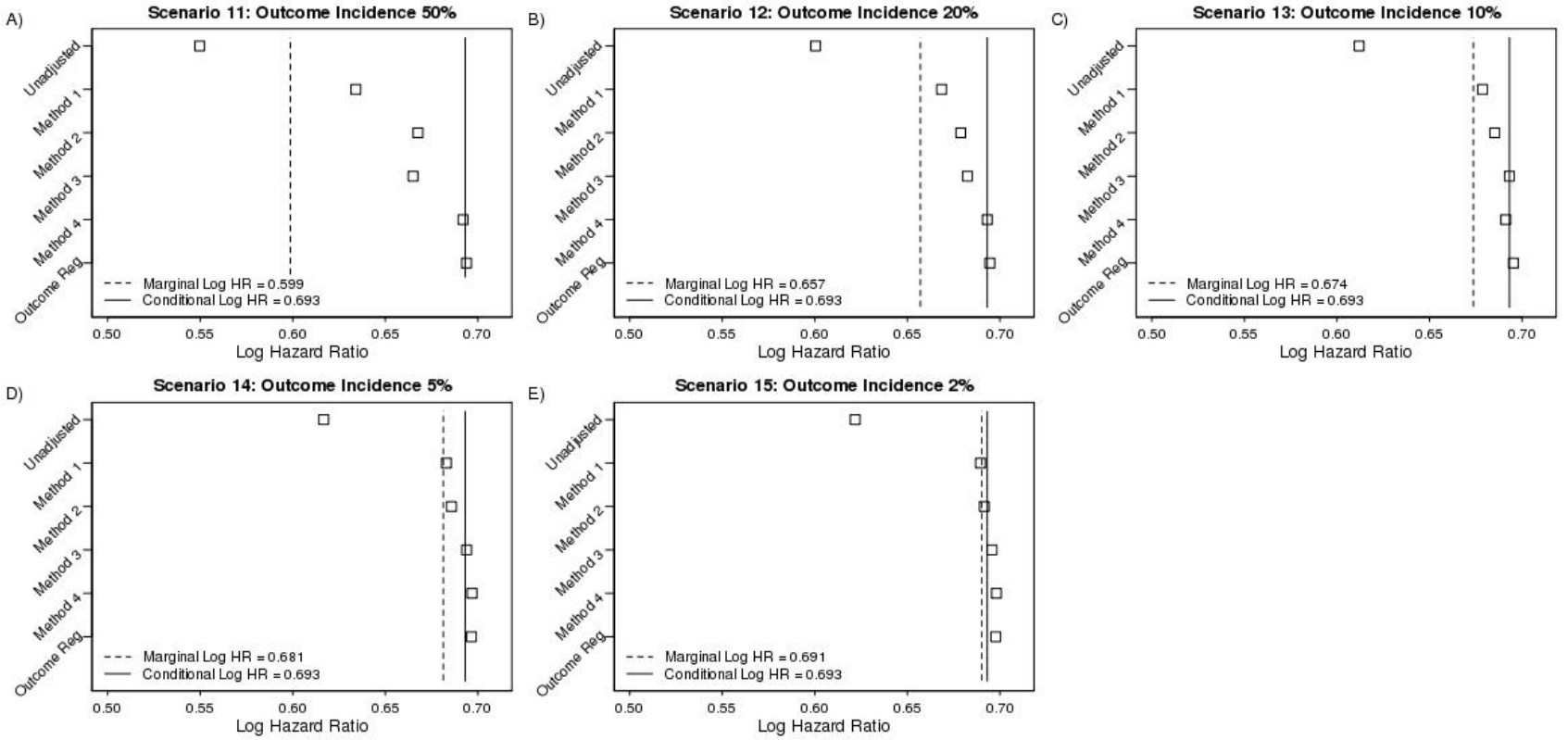
Kaiser Permanente Northern CA

- Bruce Fireman

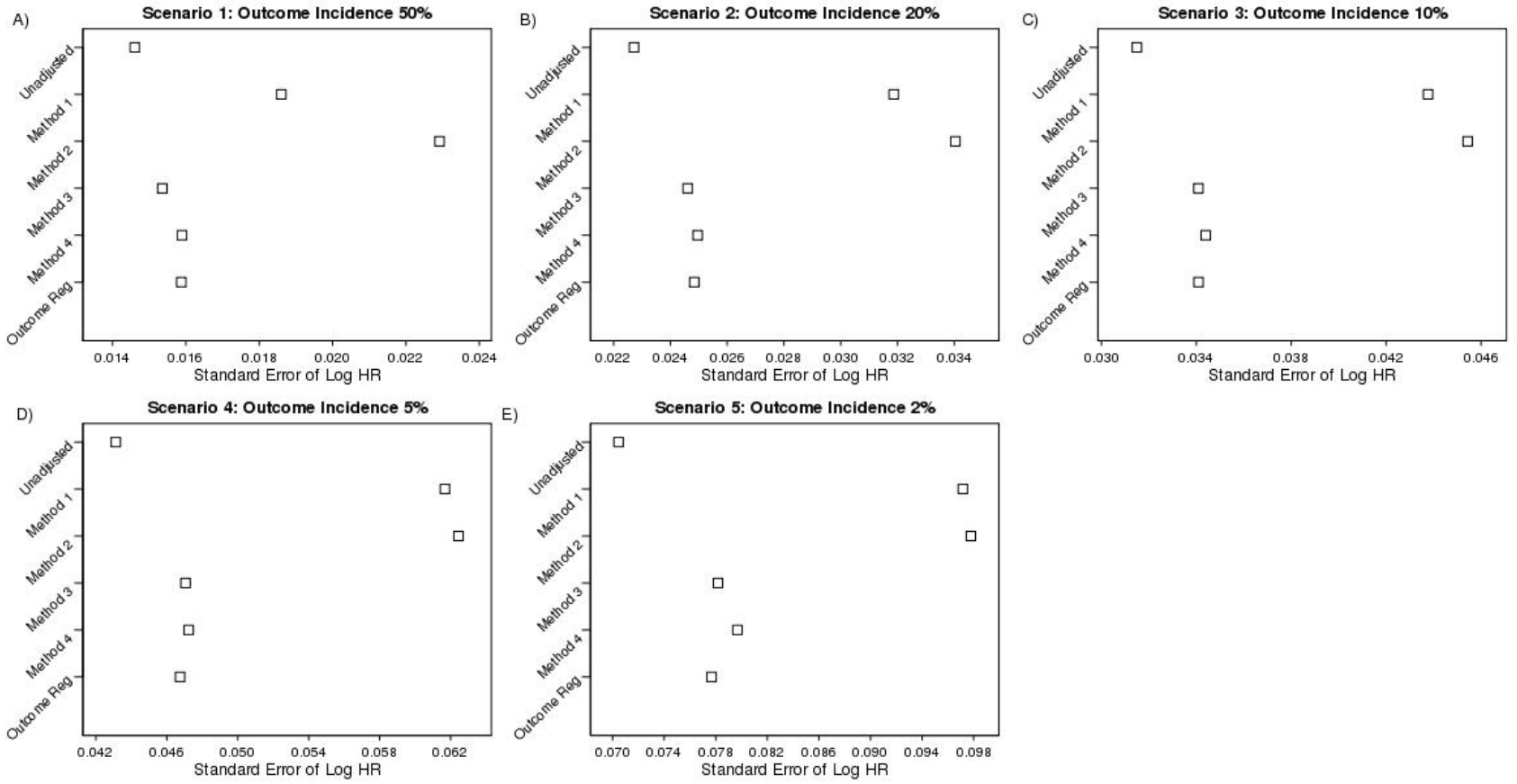
Harvard Pilgrim Healthcare Inst.

- Margaret Johnson
- Darren Toh

Thank you!



Supplemental Figure 1 (Same as Figure 5, but with censoring). Estimated log hazard ratios for Scenarios 11 through 15 with censoring before the end of follow-up. Treatment effects were estimated using 1-to-1 matching on the baseline propensity score (Method 1), 1-to-1 matching on the baseline propensity score with stratification on matched sets (Method 2), including quadratic splines of the baseline propensity score in the Cox outcome model (Method 3), including quadratic splines of the time-specific propensity scores in the Cox outcome model (Method 4), and outcome regression that adjusted for each of the baseline covariates directly in the Cox outcome model.



Supplemental Figure 2. Standard deviation of the sampling distribution of the estimated log hazard ratios across all simulation runs for Scenarios 1 through 5 with no censoring before the end of follow-up. Treatment effects were estimated using 1-to-1 matching on the baseline propensity score (Method 1), 1-to-1 matching on the baseline propensity score with stratification on matched sets (Method 2), including quadratic splines of the baseline propensity score in the Cox outcome model (Method 3), including quadratic splines of the time-specific propensity scores in the Cox outcome model (Method 4), and outcome regression that adjusted for each of the baseline covariates directly in the Cox outcome model.