EPID 765

Pharmacoepidemiology

Disease Risk Scores
(slides adapted from Richie Wyss)

1

Disease Risk Scores

- Predicted probability of outcome (w/o rx)
  - Potential advantages vs. PS:
    - Biologic (PS driven by non-biologic factors)
      - More stable over time
      - More stable across populations (cave: coding changes/diff)
      - Can be estimated prior to 1st patient being treated!
    - Meaningful scale for treatment effect heterogeneity
  - Disadvantages:
    - Does not lead to covariate balance across rx cohorts
    - Cannot evaluate balance within entire population
    - Can be difficult to estimate

2

DRS Estimation Strategies

- DRS formally defined as $E[Y_0|X]$ (Hansen 2008)
- Y_0 potential outcome had individual received no treatment
- Can only be estimated for untreated (or comparator)
- Cannot be estimated directly for treated individuals

Subject ID	Treatment Status	Y_1	Y_0	Observed Outcome (Y)
164	1	1	?	1
165	0	?	0	0
166	0	?	1	1
167	1	0	?	0

3

Traditional DRS Estimation Strategy

- Same sample DRS estimation

1. Fit outcome model within full cohort with term for treatment, predict DRS for each individual setting treatment status to untreated (Miettinen 1976)
   Assumes no treatment effect heterogeneity!
2. Fit outcome model within untreated only (restrict to untreated individuals)
   Allows for heterogeneity, but DRS prediction better in untreated used to estimate DRS (overfitting)

4

Out-of-Sample DRS Estimation

- Avoids problems of same sample DRS estimation (Hanson 08 – previous slide!)
- E.g., historical data prior to treatment introduction (Glynn et al. 12)
  - Potential additional advantages
    - Ample data for fitting rich DRS model
  - Requires additional assumptions
    - E.g., covariate effects on the outcome, coding practices, indication, and surveillance of individuals don't change over time (probably not exhaustive)

5

Baseline Covariates (selected)

Demographics:	Warfarin	Dabigatran
Age	79.01	76.70
Race (1 white, 0 other) (%)	89.28	91.94
Sex (% female)	40.85	46.76
Diagnoses (%)		
Cardiovascular:		
Chest pain	42.90	41.41
Heart disease	77.36	69.67
Heart failure	33.79	21.93
Hypertension	65.27	63.23
Prior Myocardial Infarction	4.12	2.45
Cerebrovascular disease	23.01	19.19
Prior Stroke	7.33	5.56
VTE	10.29	1.93
Diabetes	35.14	30.06
Kidney disease	13.00	4.78

6

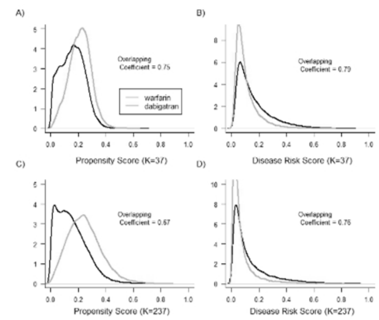
PS and DRS Estimation

- Considered 2 historical DRS models
  - Reduced model:
    - 37 a priori selected covariates
  - High dimensional model
    - 200 empirically selected covariates + 37 a priori covariates
    - Used algorithm similar to HDPS (simplified)
      - Identified top n most prevalent codes from each data dimension (e.g., inpatient/outpatient diagnoses, medication claims, etc.)
      - Included top 200 codes based on the strength of the univariate association between each code and the outcome
- Fit 2 PS models controlling for the same set of covariates for comparison
- 20% and 1% samples of ffs Medicare

7

Different Overlap PS and DRS 20%

covariates: 37 / 237



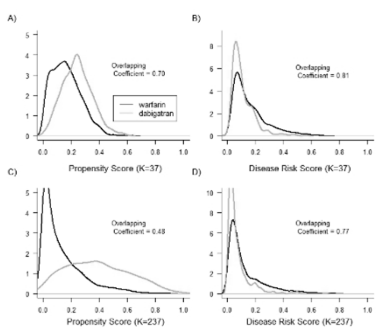
PS DRS

Initiators of dabigatran (gr) and warfarin (bl) in Medicare October 2010 to December 2012
Wyss et al. PDS 2015

8

Different Overlap PS and DRS 1%

covariates: 37 / 237



PS DRS

Initiators of dabigatran (gr) and warfarin (bl) in Medicare October 2010 to December 2012
Wyss et al. PDS 2015

9

Table 3. Empirical results comparing new users of dabigatran with new users of warfarin in preventing combined ischemic stroke and all-cause mortality in the Medicare population between October 19, 2010 and December 31, 2012.

Sample Size[a]	# covs	Method	Hazard Ratio[b]	St. Error	95% CI	% matched	Model Fit c-stat	Model Fit p-value	Model Fit ASAMD
20% Sample									
		Unadjusted	0.48	0.02	(0.46, 0.50)	------	------	------	0.14
	37								
		PS match	0.73	0.03	(0.69, 0.77)	100	0.68	0.63	<0.01
		DRS match	0.72	0.03	(0.68, 0.76)	100	0.73	<0.01	------
	237								
		PS match	0.88	0.04	(0.81, 0.95)	100	0.73	0.52	<0.01
		DRS match	0.87	0.04	(0.80, 0.94)	100	0.78	<0.01	------
1% Sample									
		Unadjusted	0.47	0.07	(0.41, 0.54)				0.17
	37								
		PS match	0.75	0.16	(0.55, 1.03)	98.3	0.71	0.65	0.01
		DRS match	0.75	0.15	(0.56, 1.01)	100	0.73	0.18	------
	237								
		PS match	0.89	0.21	(0.59, 1.34)	81.5	0.79	0.61	0.01
		DRS match	0.86	0.18	(0.60, 1.22)	99.3	0.77	<0.01	------

[a] 20% (N=67,667) and 1% (N=3,383) samples of the Medicare data.
[b] RELY trial relative risk for 150mg dabigatran vs warfarin: 0.76 (0.60, 0.98) for ischemic stroke; 0.88 (0.77, 1.00) for death from any cause. In the current study, >90% of the outcomes were death from any cause.

Wyss R, Ellis AR, Brookhart MA, Jonsson Funk M, Girman CJ, Simpson R, Stürmer T. Matching on the disease risk score in comparative effectiveness research of new treatments. Pharmacoepidemiology and Drug Safety 2015;24(9):951-61.

10

Covariate Balance PS vs DRS

- PS
  - Check balance of covariates across treatment groups to assess validity of PS model
    - Strong correspondence between covariate balance and ability of PS model to control confounding (Franklin et al. 2014, Ali et al. 2014)
    - Cave: could still be imbalanced in subgroups!
- DRS
  - "Prognostic balance" cannot be evaluated within the full study population

11

"Dry Run Analysis"

- Hansen proposes a resampling method to validate DRS models
- Resampling methods:
  - Validate models using random subsets (e.g., cross validation)
- Hansen's "Dry Run" analysis:
  - Create "pseudo treatment" group by sampling untreated in a way to represent the covariate distributions of the treated
  - Estimate "treatment" effect in the pseudo population controlling for DRS
  - Truth=no treatment effect (since no one treated)

12

Simulation Study

- Dichotomous treatment and outcome
- 6 binary covariates, 4 standard normal covariates
- Treatment and outcome model included main effects + 3 interaction terms + 2 quadratic terms
- Considered 50 different parameter combinations x 6 settings for a total of 300 unique scenarios
- For each of the 300 scenarios, we fit 32 different DRS models with different degrees of misspecification
- Correlation between 6 metrics for evaluating risk models and bias in the treatment effect
  - C-statistic
  - MSE
  - p-value from Hosmer-Lemeshow goodness of fit test
  - Pseudo bias (3 different pseudo populations)

Wyss R, Hansen BB, Ellis AR, Gagne JJ, Desai RJ, Glynn RJ, Stürmer T. Evaluating the Validity of Disease Risk Scores for Confounding Control in Non-Experimental Studies: the "Dry-Run" Analysis. American Journal of Epidemiology 2017;185(9):842-52.

13

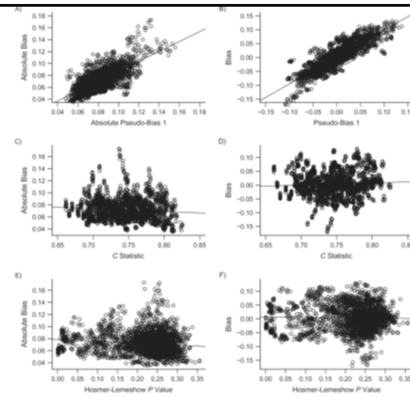
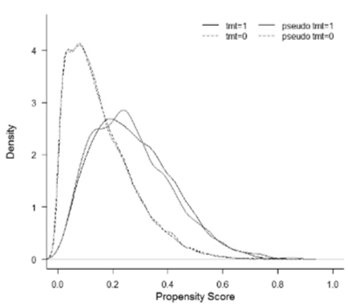


Figure 3. Metrics for evaluating DRS models plotted against the absolute bias((A), (C), and (E)) and bias ((B), (D), and (F)) in the effect estimate. Each point represents a value for the given metric (e.g., *C* statistic) and the corresponding bias in the treatment effect estimate. Each subplot contains values across all parameter combinations for each of the 32 fitted DRS models. The actual bias in the treatment effect estimate and pseudo-bias were calculated after matching the estimated disease risk scores. Pseudo-bias 1 in (A) corresponds to PS model 1 (including all higher-order terms). The line in each plot is the fitted least squares regression.

14

Empirical Study



15

Empirical Study

Table 3. Empirical results comparing dabigatran vs warfarin in the Medicare population between October 19, 2010 through December 31, 2013 (n=67,667)

# covs[a]	Method	HR (95% CI)	Pseudo bias[a]	ASAMD[b]	c-statistic	Hosmer lemeshow test[c]
237						
	Unadjusted	0.48 (0.46, 0.50)	0.45	0.12	-----	-----
	PS match	0.88 (0.81, 0.95)	------	<0.01	0.73	p=0.52
	DRS match	0.87 (0.80, 0.94)	0.01	------	0.78	p<0.01

[a] PS and DRS models included 200 empirically selected covariates and 37 covariates selected a priori.

16

Dry-Run Conclusions

- Accurately modeling the DRS within the study cohort, or within a historical set of controls presents unique challenges that are not shared by the PS
- Measures of predictive performance do not necessarily identify the ability of a DRS model to control confounding
- If the PS can be accurately modeled, evaluating the ability of the DRS model to control confounding within a "dry run" analysis provides insight into validity of fitted DRS models
  - Why not just use PS?
  - DRS can be beneficial (overlap, evaluating treatment effect heterogeneity)

17

Refined View DRS vs PS

- DRS reduces chances to condition on instruments
- DRS does not require covariate balance: compare larger proportion of individuals across treatment groups
  - Overlap in DRS distributions across treatment groups always at least as large as in PS distributions (Wyss et al. PDS 2015)
- PS model may be more complicated, but
  - Dichotomous covariates limit complexity of functional form
  - Time specific PS (Seeger et al. 2005, 2011, Mack et al. 2013)
  - Possibility for covariate balance check
- PS model may be more difficult to fit for new drugs, but
  - Overfitting PS does not necessarily compromise confounding control (Rassen et al. 11) but can reduce precision (Crowson et al. 13)

18

Conclusions DRS

- Controlling for many covariates using summary scores can improve confounding control
  - Note: recent pubs on coarsened exact matching!
- DRS may have specific advantages over PS for small samples, including newly marketed drugs
  - Increased separation with PS due to overfitting
  - Ability to estimate DRS prior to marketing of new drug
- Accurately modeling the DRS can be challenging compared to the PS, even in settings involving newly introduced treatments
  - DRS does not lead to exchangeability and can therefore not be evaluated by covariate balance!

19