Quantitative Safety & Epidemiology / CMO & PS

Propensity Score methods for causal inference in observational studies of rare outcomes

Sigrid Behr

3rd Basel Epidemiology Seminar

August 17, 2018

NOVARTIS

Acknowledgement: This presentation is based on work by my colleague Emil Scosyrev, Novartis

The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of Novartis or any of its officers.

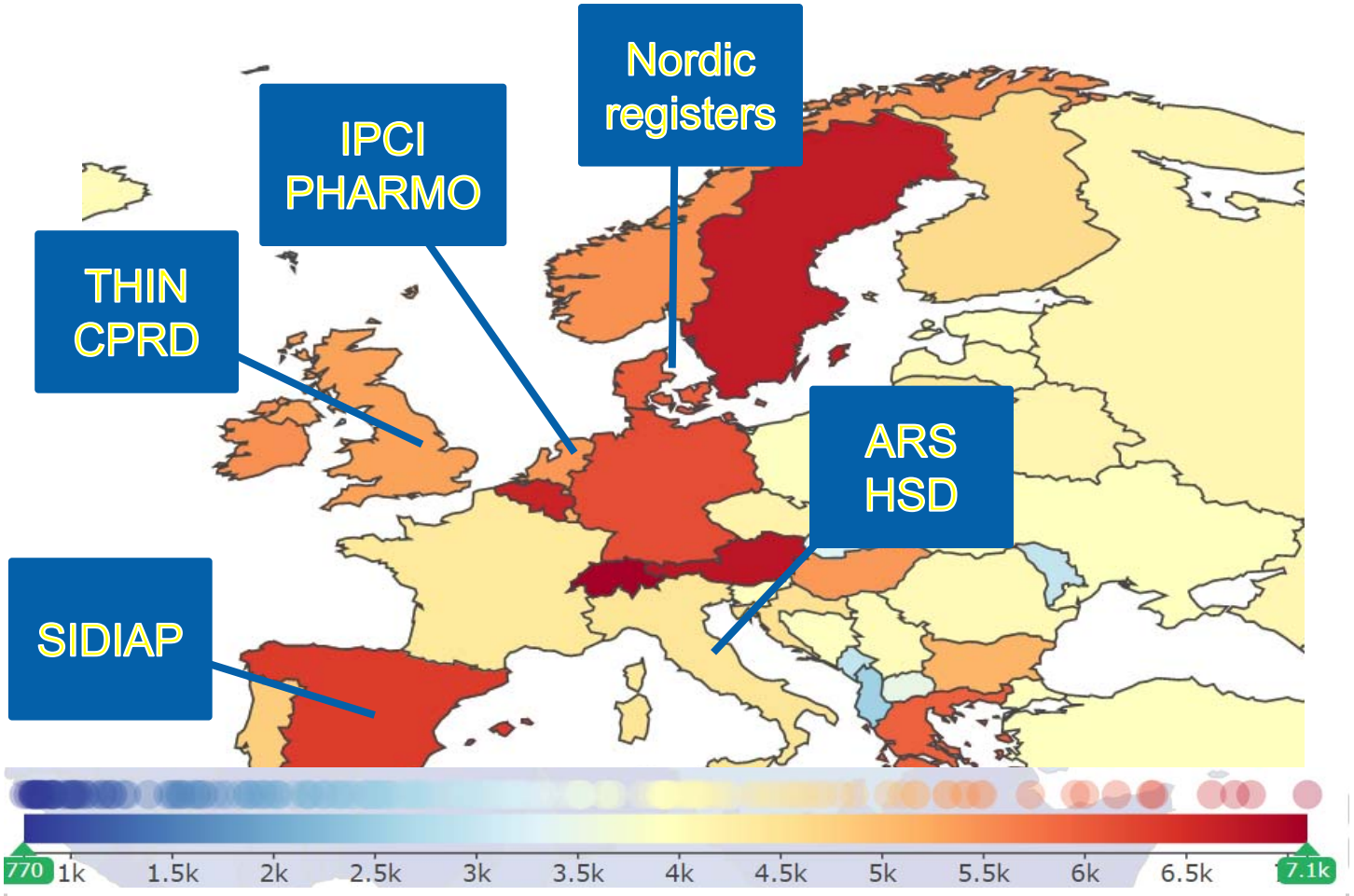
Motivation

- Remaining unknowns after clinical development program: treatment effects on rare[1)] outcomes
- Examples for rare safety outcomes of interest:
  - All-cause mortality in EU: ~8/1,000 person-years at age 50-69 (GBD 2016)
  - Stroke, myocardial infarction (COPD patients): ~1/100 person-years (Jara 2012)
  - Angioedema: ~4/1,000 person-years (Toh 2012)
- Need for large post authorization safety studies (PASS)

Incidence rate	Study size	Expected number of cases in 1 year
1/100 PY	10,000	100
1/1,000 PY	100,000	100
1/10,000 PY	100,000	10

[1)] CIOMS Working Group III frequency categories (events/persons): Uncommon: ≥ 1/1,000 – <1/100, Rare: ≥ 1/10,000 – <1/1,000, Very rare: < 1/10,000

Case study: EU multi-database approach in COPD patients



Prevalence of COPD in Europe (per 100,000), Source: GBD 2016

NOVARTIS

Case study: simplistic* setting

- Aim: compare new COPD treatment with 1 control (standard of care) regarding risk (hazard ratios) of cardiovascular events and all-cause mortality
- New user cohort study in 5 databases:
  Source population ~14 million → 700-800K COPD patients → 6,000 initiators of new treatment (planned)
- High likelihood of confounding:
  - new treatment channelled to specific patients:
    less severe disease due to physician's concerns OR more severe disease because only patients who failed on standard of care are switched
  - stage of disease correlated with occurrence of outcomes
- Approx. **50** measured potential confounders, e.g., age, sex, BMI, smoking, COPD severity, COPD duration, recent acute cardiac events, relevant medical history (diseases and treatment)

*Note: Realistic study complicated by multiple comparators, frequent treatment changes, partially missing confounder information, etc.

Case study: Analysis options - 1

How to estimate the treatment effect?

- General strategy: meta-analytic approach to account for differences in databases
- Idea 1: multivariable Cox regression analysis adjusted for all 50 potential confounders
  - need >500 events according to rule of thumb (10 events per covariate)
  - Or apply covariate selection algorithm → BUT: covariate selection algorithms affected by small data problems like invalid p-values and finite sample bias

Case study: Analysis options - 2

How to estimate the treatment effect?

- Idea 2: Propensity Score (PS) methods to adjust for confounders in rare outcome situation
  - Estimate probability of receiving new treatment given the observed covariates by logistic regression model in each database
  - Challenge: only ~170 «treatment events» expected in smallest database → small sample problem also for PS model (3.4 events/covariate)?
- How to use the PS?

Case study: Analysis options - 3

How to estimate the treatment effect with PS methods?

Options:

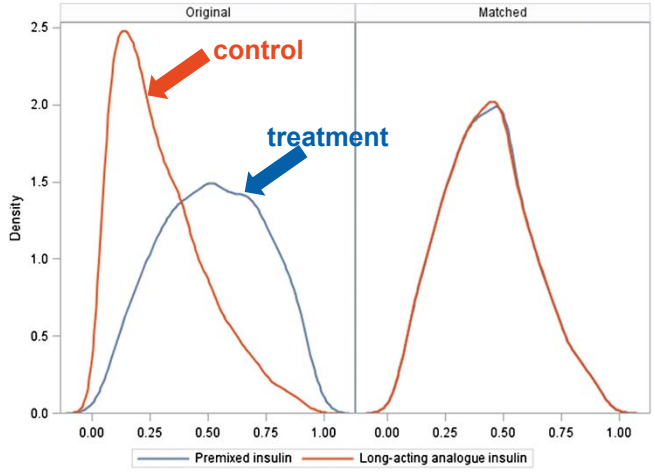
A. PS matching: 1. match treated and control patients on PS, 2. Cox regression on matched cohort

B. PS stratification: Cox regression within strata of PS

C. PS adjustment: PS as covariate in outcome regression

D. Inverse Probability of Treatment Weighting (IPTW)

PS methods – A: PS matching

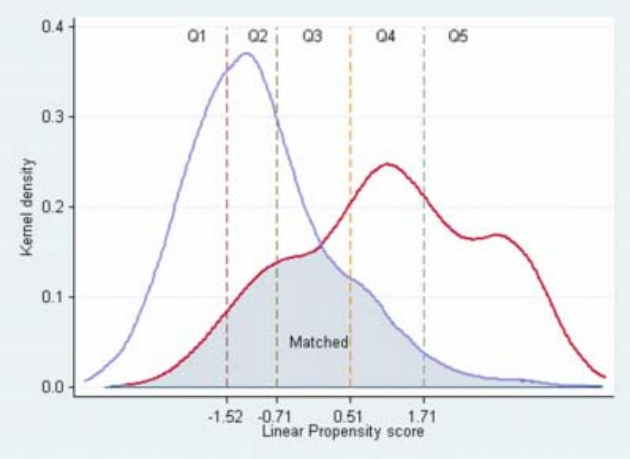


- Strategy: match n control subjects with similar PS to each treated subject
- What happens:
  - Patients with PS≈1 (clear indication) excluded due to lack of matching partners
  - Patients with PS≈0 («contra-indication») excluded due to lack of matching partners

Adapted from: Kollhorst et al. 2015, https://doi.org/10.1111/dom.12554

- Expected consequences for analysis:
  - Efficiency loss due to removal of patients → larger standard errors, wider confidence intervals
  - If sufficient overlap of PS curves, matched cohort characteristics similar to characteristics of treated population → causal contrast estimated in outcome regression: effect of treatment on the treated

PS methods – B: PS stratification



Wang, 2008. The Internet Journal of Epidemiology, 7(2)

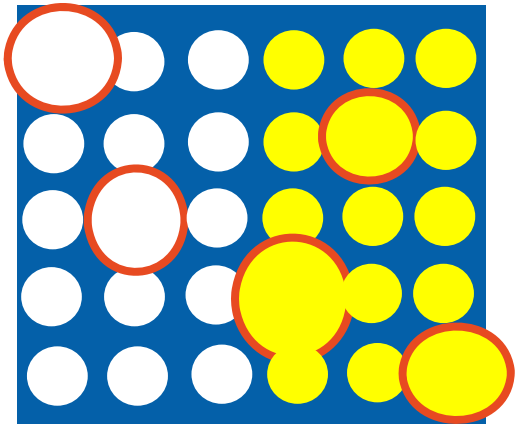
- Strategy: build strata based on quantiles of the joint PS distribution
- What happens:
  - Full cohort used in analysis
  - First and last quantile may include wide range of PS

- Expected consequences for analysis:
  - More efficient than matched analysis
  - Residual confounding may happen; finer stratification could mitigate risk of residual confounding but re-introduces small sample problems
  - Causal contrast estimated in outcome regression: average treatment effect in population

PS methods – C: PS adjustment

- Strategy: Include PS as continuous covariate in outcome regression
- More questions than answers:
  - How to specify the model? Is a simple linear model sufficient or do we need more complex functional forms?
  - What happens if the functional form is misspecified?
  - What exactly does the treatment effect mean?

PS methods: – D: IPTW 1



- IPTW idea: create a pseudo-population in which treatment and covariates are independent

- Step 1: calculate weights for each person i:

$$w_i(x) = \begin{cases} \frac{1}{PS_i}, & if\ person\ i\ received\ new\ treatment \\ \frac{1}{1-PS_i}, & if\ person\ i\ received\ control \end{cases}$$

- Step 2: outcome regression on weighted population (e.g. Cox regression using a robust variance estimator)

PS methods: – D: IPTW 2

Example adapted from: Robins et al. 2000

Conf.	Treatm.	Outcome	N Obs. Population		PS	weight	N Pseudo-Population	
1	1	1	108	360	0.9	1.11	120	400
1	1	0	252		0.9	1.11	280	
1	0	1	24	40	0.1	10	240	400
1	0	0	16		0.1	10	160	
0	1	1	20	50	0.5	2	40	100
0	1	0	30		0.5	2	60	
0	0	1	40	50	0.5	2	80	100
0	0	0	10		0.5	2	20	

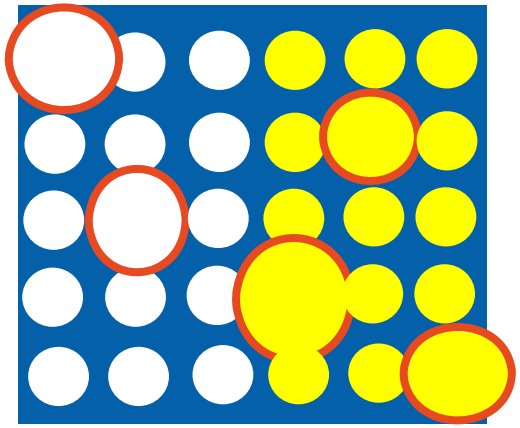
PS methods: – D: IPTW 3

- Problem with weights: large weights lead to instable results
- Solution: use **stabilized** weights:

$$w_{i,S}(x) = \begin{cases} \dfrac{N_T}{N}\dfrac{1}{PS_i}, & if\ person\ i\ received\ new\ treatment \\ \dfrac{N_C}{N}\dfrac{1}{1-PS_i}, & if\ person\ i\ received\ control \end{cases}$$

$\frac{N_T}{N}$, $\frac{N_C}{N}$: baseline probability of receiving treatment or control

PS methods: – D: IPTW 4



- Strategy: Apply IPTW with stabilized weights
- What happens:
  - Full cohort used in analysis
  - Patients treated against expectation (PS≈0 among treated or PS≈1 among controls) may still lead to large weights

- Expected consequences for analysis:
  - More efficient than matched analysis
  - Pseudo-population balanced for observed covariates (can be checked e.g. by plotting the standardized differences)
  - Extreme weights may influence analysis → trim extreme weights, i.e. remove these patients

Case study: Analysis options - 4

Assumed advantages and disadvantages:

Analysis method	+	-
Multivariable regression	No black-box model, can see how covariates behave (risk factor or protective)	High likelihood of bias due to small sample problems in model or covariate selection
PS matching	Concept intuitive, easy to assess covariate balance after matching	Low efficiency due to exclusion of subjects If lack of overlap, unclear which treatment effect is estimated
PS stratification	Easy to perform, all subjects included	Potential for residual confounding; assessment of covariate balance within strata possible but inconvenient
PS adjustment	?	Potential bias if functional form misspecified? Difficult to assess if PS worked
IPTW	Probably efficient and unbiased, if no extreme weights	Pseudo-population approach not intuitive (artificial population)

What works best in practice?

➢Conduct simulation study for this specific setting

NOVARTIS

Simulation study – setting 1

2,000 replications of the following cohort study:

- Data generation:
  - Cohort: N= 1,000 patients
  - 50 covariates: X_1- X_{40} binary (0/1), X_{41}- X_{50} continuous covariates
  - Outcome depends on all 50 covariates
  - Treatment assignment depends on X_1- X_5 and X_{41}- X_{45} (i.e. only these are confounders):
    - generated by logistic function with slope ln(1.2) and intercept -2
    - i.e., odds of treatment increased by 1.2 per unit increase in covariate
    - results in treatment prevalence of 17%
  - Outcome variable: exponential model with intercept -5.25, slope ln(1.2) for covariates, treatment hazard ratio=2 → results in approx. 50 events

Simulation study – setting 2

- Analysis methods:
  - Cox regression
    - Univariable
    - Multivariable: with all 50 covariates, using stepwise selection
  - PS methods (PS model with all 50 covariates or using stepwise selection):
    - Matching: 1:1 matching on PS deciles[1)]
    - Stratification: on PS deciles
    - Adjustment: PS as continous covariate assuming log-linear association with outcome
    - IPTW: using stabilized weights
- For comparison: random treatment assignment & univariable Cox regression (as in randomized controlled trial -RCT)

[1)] Smaller caliper would have resulted in exclusion of treated subjects

Simulation study – measures of interest 1

Bias

- $\exp(\hat{\theta})$ = estimated hazard ratio for the treatment effect (mean over 2000 replications)

Efficiency

- $\widehat{SE}$ = estimated standard error of $\hat{\theta}$ from the Cox model (mean)
- $\widehat{SE}/\widetilde{SE}$ as measure of bias for $\widehat{SE}$, where $\widetilde{SE}$ = empirical standard error of $\hat{\theta}$ in 2000 replications

Coverage

- Proportion of 95% CIs capturing the true hazard ratio

Simulation error

- {LCL, UCL} = lower and upper 95% confidence limits representing simulation error

Simulation study – measures of interest 2

Bias	• $\exp(\hat{\theta})$ ≈ 2 means minimal bias
Efficiency	• $\widehat{SE}$: the smaller the better as long as $\widehat{SE}/\widetilde{SE}$≥1 • $\widehat{SE}/\widetilde{SE}$≈ 1; $\widehat{SE}/\widetilde{SE}$<1 indicates underestimation of $\widehat{SE}$
Coverage	• Should be close to 95%
Simulation error	• Width of interval

Simulation study results - 1

Cox Model	Mean $\exp(\hat{\theta})$	95% LCL	95% UCL	95% CI Coverage	95% LCL	95% UCL	Mean $\widehat{SE}(\hat{\theta})$	95% LCL	95% UCL	$\frac{\widehat{SE}(\hat{\theta})}{\widehat{SE}(\hat{\theta})}$
1. Univariable crude	2.933	2.894	2.972	0.758	0.739	0.776	0.291	0.290	0.292	0.965
2. Multivariable-50	2.369	2.329	2.411	0.916	0.904	0.928	0.363	0.361	0.365	0.919
3. Multivariable-step	2.326	2.291	2.362	0.900	0.886	0.913	0.302	0.301	0.304	0.874
4. Matched PS-50	2.274	2.233	2.317	0.958	0.949	0.966	0.408	0.405	0.410	0.972
5. Matched PS-step	2.361	2.318	2.405	0.964	0.956	0.972	0.411	0.408	0.413	0.975
6. Stratified PS-50	2.113	2.083	2.142	0.947	0.937	0.957	0.309	0.308	0.311	0.975
7. Stratified PS-step	2.189	2.159	2.219	0.940	0.929	0.950	0.302	0.301	0.303	0.968
8. Continuous PS-50	2.081	2.052	2.111	0.949	0.939	0.959	0.320	0.319	0.322	0.990
9. Continuous PS-step	2.168	2.138	2.198	0.945	0.935	0.955	0.310	0.308	0.311	0.972
10. IPTW-50	2.021	1.989	2.053	0.944	0.934	0.954	0.345	0.343	0.347	0.959
11. IPTW-step	2.111	2.080	2.143	0.942	0.932	0.952	0.324	0.322	0.325	0.959
12. RCT model	2.022	1.994	2.050	0.961	0.952	0.969	0.318	0.317	0.320	0.992

Simulation study - discussion

- Using stepwise covariate selection algorithm results in higher bias
- PS methods performed better than univariable or multivariable Cox regression:
  - PS matching least efficient
  - IPTW resulted in low bias and acceptable efficiency in our simulation; another simulation study (Franklin 2017) found IPTW biased if tails of PS distributions do not overlap
  - PS adjustment showed good performance in simulations, however, simulations did not probe impact of misspecification of functional form → theoretical concerns remain

Conclusions

Analysis strategy for case study:

- PS method: IPTW
- (+ univariable and multivariable Cox regression adjusted for a priori defined key confounders)

Recommendations for rare outcome, uncommon treatment situation with many potential confounders:

- Avoid covariate selection algorithms
- PS methods outperform multivariable regression

References

- CIOMS Working Group III, 1995: Preparing Core Clinical Safety Information on Drugs. ISBN: 9290360623
- Franklin et al. Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. Statistics in Medicine 2017; 36
- Global Burden of Disease Collaborative Network.Global Burden of Disease Study 2016 (GBD 2016) Results.Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.Available from http://ghdx.healthdata.org/gbd-results-tool.
- Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington, 2016. Available from http:// vizhub.healthdata.org/gbd-compare. (Accessed 04Aug2018)
- Jara et al. A new user cohort study comparing the safety of long-acting inhaled bronchodilators in COPD. BMJ Open 2012; 2:e000841.
- Kollhorst et al. Comparison of basal insulin therapies with regard to the risk of acute myocardial infarction in patients with type 2 diabetes: an observational cohort study. 2015 https://onlinelibrary.wiley.com/doi/abs/10.1111/dom.12554
- Robins et al. Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology 2000; 11(5).
- Toh et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Arch Intern Med. 2012; 172(20)
- Wang. Propensity score methods to adjust for confounding in assessing treatment effects: bias and precision. The Internet Journal of Epidemiology. 2008 Volume 7 Number 2.

Thank you

NOVARTIS

Details of simulation study: data generation

- X1 ~ $B(0.1)$
- $X2 = 1\ if\ X1 = 1\ and\ X2 \sim B(0.2)\ if\ X1 = 0$
- $X3 \sim B(0.3)$
- $X4 = 1\ if\ X3 = 1\ and\ X4 \sim B(0.4)\ if\ X3 = 0$
- $X5 \sim B(0.5)$
- $X6 \sim B(0.06), X7 \sim B(0.07), \dots, X40 \sim B(0.40)$

- $X41 \sim N(0,1)$
- $X42 = X41 + N(0,1)$
- $X43 \sim N(0,1)$
- $X44 = X43 + N(0,1)$
- $X45 \sim N(0,1)$
- $X46, \dots, X50 \sim N(0,1)$

Where B(m) is the binomial distribution with success probability m and $N(0,1)$ is the standard normal distribution with mean 0 and variance 1.