

# Anforderungen an die Daten: Eine Diskussion anhand von Patientenregistern

## Workshop Praktische Aspekte bei der Anwendung von Propensity Scores

Tim Mathes

Institute für Medizinische Statistik

AG Clinical Epidemiology and Health Economics

e-mail: [tim.mathes@med.uni-goettingen.de](mailto:tim.mathes@med.uni-goettingen.de)

# What is a registry in health care?

- No consistent definition:
  - A systematic collection of data
  - An organized system that uses observational study methods to collect uniform data (...) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure (...)
- A type of systematic collected “Real-World-Data” (versorgungsnahе Daten)
- Observational data!



# Registries for HTA

rbNRS should be based on causal inference methods, which aim to emulate the effect estimate of a RCT (emulation of target trial concept)

→ Estimand of interest in HTA:

- Treatment Policy (Intention-To-Treat-Analyses)
- Average Treatment Effect in the total population

# Potential source of bias/variation

## ORIGINAL ARTICLE

No inexplicable disagreements between real-world data—based nonrandomized controlled studies and randomized controlled trials were found

- Most effect estimates agreed
- In the case there were relevant differences this usually could have been foreseen and either probably could have been avoided (e.g. inadequate analysis) or should have permitted the conduct of an rbNRS (e.g. important confounders not in the database)
- In none of the studies a true treatment policy estimand was calculated and the estimand was often unclear or undefinable
- Side note: potential drivers for the effectiveness-efficacy gap appeared to have little impact

# Source of bias in rbNRS (unpublished project)

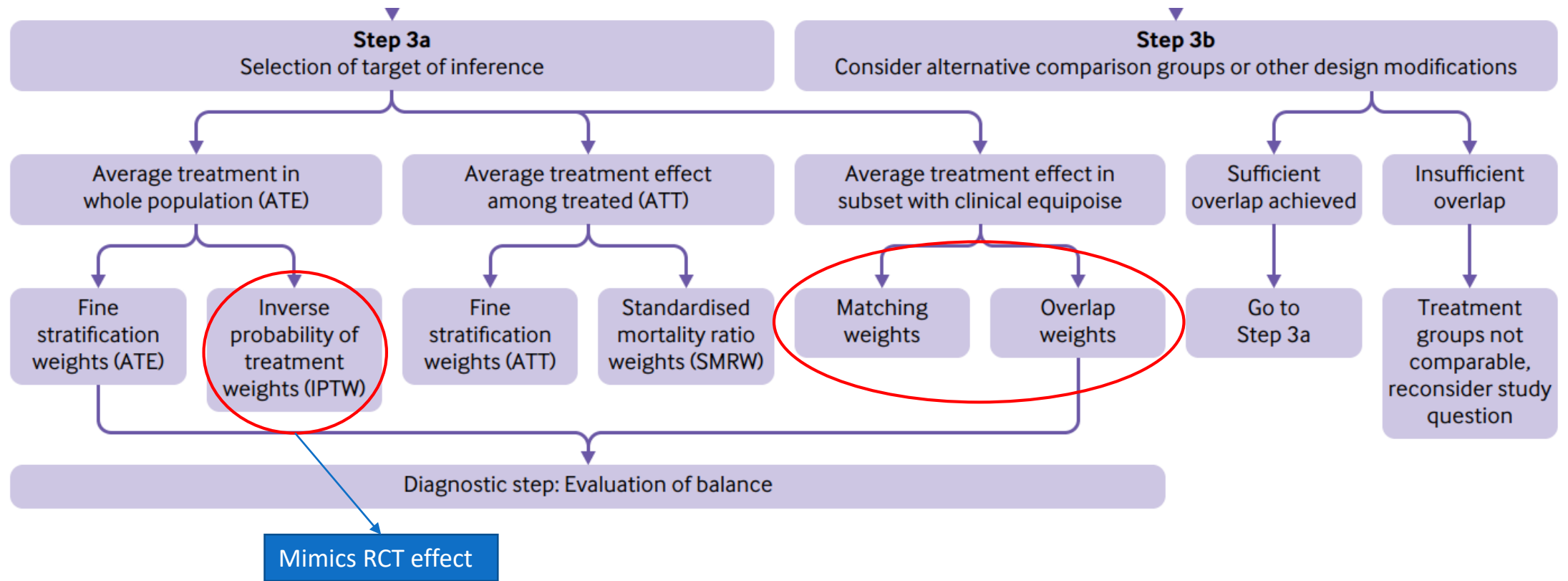
- Confounding
  - Most analyses did not incorporate all relevant potential confounders
  - or measurement of incorporated confounding variables was unclear
- Selection
  - For many studies there was a risk of immortal time bias
- Classification
  - Some studies were at risk for classification bias (more concrete linkage bias)
- Deviations from intended interventions
  - None of the studies estimated a true ITT effect and a clear estimand could not be determined (“starting and ???”)
- Missing data
  - In none of the studies missing data could be sufficiently assessed

Most problem were data related but searching and checking primary registry reports does not improve the situation relevantly

# Summary problems

*„Without clear guidance on the connection between the research question, available data, and assumptions and properties of different causal inference methods, researchers often apply suboptimal methods, analytical findings suffer from serious flaws, and important topics in CER go unanswered or are answered incorrectly” (PCORI 2019)*

# Selecting the Propensity Score Method



# Analysis strategies for emulation

rbNRS should be based on causal inference methods which aim to emulate the effect estimate of a RCT (emulation of target trial concept)

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán\* and James M. Robins

## Comparing the performance of propensity score methods in healthcare database studies with rare outcomes

Jessica M. Franklin,<sup>a</sup> Wesley Eddings,<sup>a</sup> Peter C. Austin,<sup>b</sup> Elizabeth A. Stuart<sup>c</sup> and Sebastian Schneeweiss<sup>a</sup>

The first example comes from a study of 18,475 patients initiating either dabigatran (the new treatment) or warfarin (control) from October 2010 (when dabigatran was first approved for sale in the USA) through

The second study included 166,031 patients who had initiated an anticonvulsant medication between 2001 and 2006 [50]. Anticonvulsant exposure was classified as 'highly inducing', meaning those that

Most methods were originally developed for Big-Data and including more confounders (that effect outcome) is usually better!



# German registries: Availability of data

- About 356 registries in Germany (2019)
- Only 29,4% generally accessible
- >50%, <2500 observations
  - Only a (small) share will be the patients, intervention/comparator of interest, in particular because abDa is mostly on rare diseases
  - Data on all (or linkage) outcomes of interest must be available
  - Data on all (only linkage or matching) confounders must be measured
  - Loss due to trimming because of non-overlapping regions
  - In case of improving balance of confounders (e.g. matching) sample size could be further reduced
- Only 36 regular quality (indicators) reports
  - Data must be of sufficient quality and quality must be accessible



# Analysis strategies for emulation

Causal inference methods for small non-randomized studies: Methods and an application in COVID-19

Sarah Friedrich\*, Tim Friede

Department of Medical Statistics, University Medical Center Göttingen

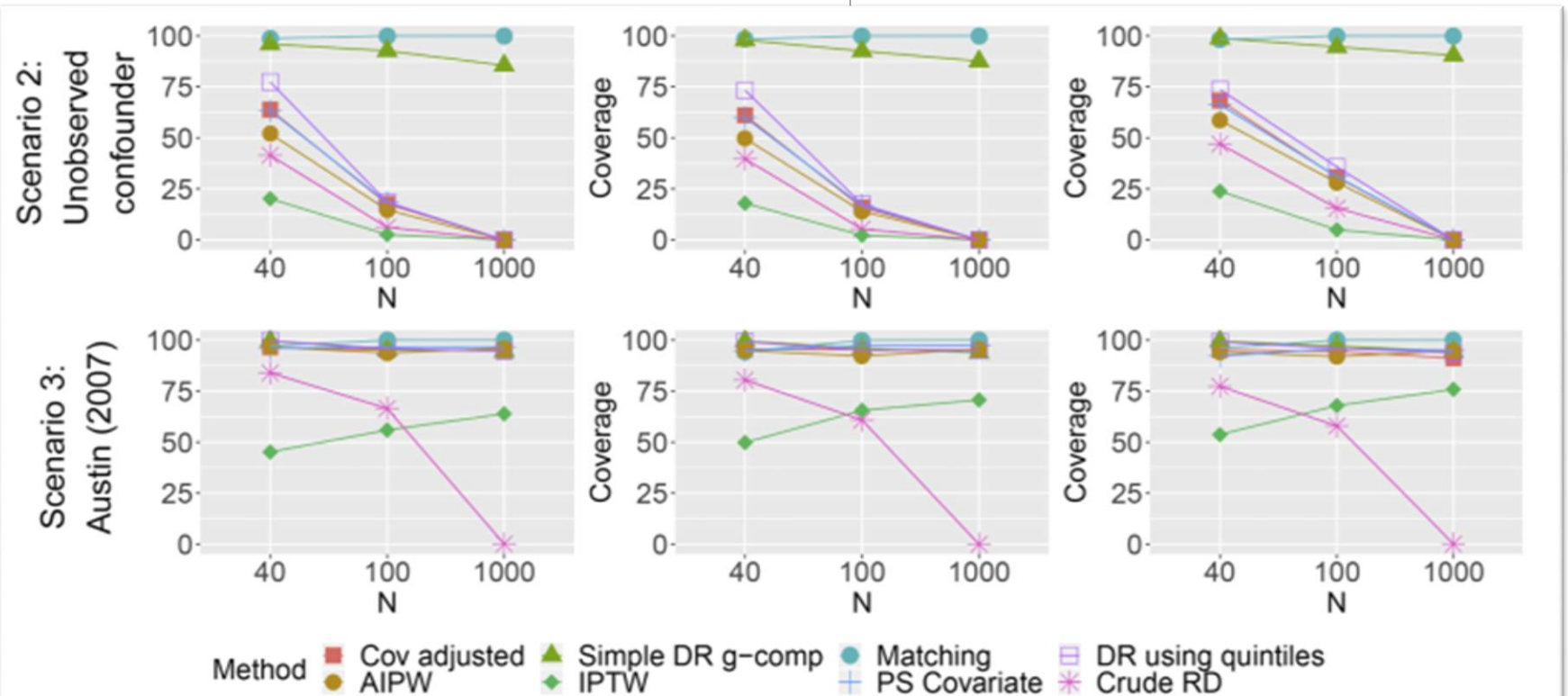


Fig. 2. Displayed is the coverage probability (in %) for the three scenarios (rows) and the three simulated risk differences (columns).

approaches. Interestingly, it turned out that the default settings in software implementations are often more suitable for large sample sizes and need to be adjusted for applications in small-scale studies. For

# Registries: Applicability of data

- >50% limited to certain centers (e.g. Universities)  
➡ Real-World but generalizable???
- PS method (e.g. Matching) and improvement of balance could have an effect on generalizability



# Results of first attempts to systematically emulate RCTs

Circulation

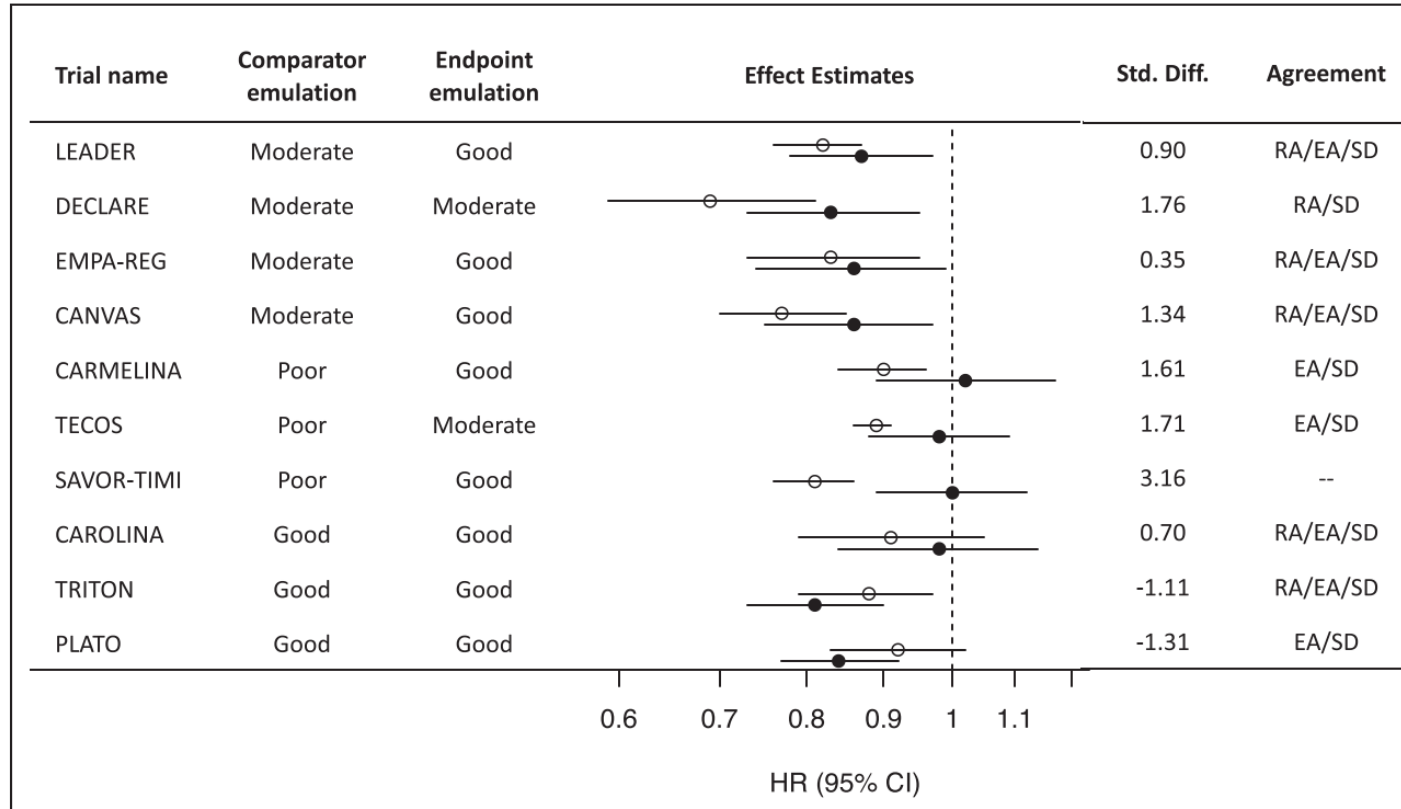
ORIGINAL RESEARCH ARTICLE

## Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies

First Results From the RCT DU

**CONCLUSIONS:** Agreement between RCT and RWE findings varies depending on which agreement metric is used. Interim findings indicate that selection of active comparator therapies with similar indications and use patterns enhances the validity of RWE. Even in the context of active comparators, concordance between RCT and RWE findings is not guaranteed, partially because trials are not



needed to understand how often and




# Results of first attempts to systematically emulate RCTs

ORIGINAL ARTICLE | VOLUME 151, P161-170, NOVEMBER 01, 2022

## Understanding variation in the results of real-world evidence studies that seem to address the same question

Shirley V. Wang   • Sushama Kattinakere Sreedhara • Lily G. Bessette • Sebastian Schneeweiss

Published: September 05, 2022 • DOI: <https://doi.org/10.1016/j.jclinepi.2022.08.012> •  Check for updates

### Results

Most variation remained unexplained (60–88%). Of the explained variation, two-thirds were related to data and population differences, and one-third were related to the use of alternative study design and analysis parameters. Among these, the most prominent were differences in outcome algorithms and criteria used to define follow-up.

### Conclusion

When making policy decisions based on database study findings, it is important to evaluate the validity, consistency, and robustness of results to alternative design and analysis decisions.

# Conclusion

- There are common data related problems when using registry data to emulate trials (immortal time bias, estimands, missing data)
  - Analysis should be carefully planned to avoid, “avoidable” mistakes
  - German registry data currently appear to have limited potential to be used for target trial emulations
  - In small samples standard methods may need to be adapted
  - Unexplainable uncertainty remains!
- ➡ Excepting uncertainty is only acceptable if the effort is relevantly lower than for a pragmatic RCT?

Thanks for your attention

# References

1. Franklin, J. M., Eddings, W., Austin, P. C., Stuart, E. A., & Schneeweiss, S. (2017). Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. *Statistics in medicine*, 36(12), 1946-1963.
2. Friedrich, S., & Friede, T. (2020). Causal inference methods for small non-randomized studies: Methods and an application in COVID-19. *Contemporary clinical trials*, 99, 106213.
3. Gliklich R, Dreyer N. Registries for evaluating patient outcomes: a user's guide, (prepared by Outcome Decide Center, Rockville [Outcome Sciences, Inc. d/b/a Outcome] under Contract No. HHS A290200500351 TO3). AHRQ Publication No; 2007.
4. Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8), 758-764.
5. Landsittel D, Chang C-C H, Morton, SC. (2019). Guidance for Researchers on Optimal Methods for Conducting Comparative Effectiveness Research With Observational Data. Washington, DC: Patient-Centered Outcomes Research Institute (PCORI).  
<https://doi.org/10.25302/11.2019.ME.130603827>
6. Mathes, T., Rombey, T., Kuss, O., & Pieper, D. (2021). No inexplicable disagreements between real-world data–based nonrandomized controlled studies and randomized controlled trials were found. *Journal of Clinical Epidemiology*, 133, 1-13.
7. Niemeyer, A., Semler, S. C., Veit, C., Hoffmann, W., Röhrig, R., Gurisch, M. S. C., ... & Beckedorf, D. P. F. I. (2021). Gutachten zur Weiterentwicklung medizinischer Register zur Verbesserung der Dateneinspeisung und-anchlussfähigkeit. Hamburg/Berlin.
8. Schneeweiss, S., Franklin, J., & Wang, S. (2021). Emulating randomized clinical trials with non-randomized real-world evidence studies.
9. Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj*, 355.
10. Wang, S. V., Sreedhara, S. K., Bessette, L. G., & Schneeweiss, S. (2022). Understanding variation in the results of real-world evidence studies that seem to address the same question. *Journal of Clinical Epidemiology*, 151, 161-170.
11. Zhang, Y., Yang, S., Ye, W., Faries, D. E., Lipkovich, I., & Kadziola, Z. (2021). Practical recommendations on double score matching for estimating causal effects. *Statistics in medicine*.