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
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” (versorgungsnahe 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

- 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Source of bias in rbNRS (unpublished project)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

Step 3a: Selection of target of inference
- Average treatment in whole population (ATE)
  - Fine stratification weights (ATE)
  - Inverse probability of treatment weights (IPTW)
- Average treatment effect among treated (ATT)
  - Fine stratification weights (ATT)
  - Standardised mortality ratio weights (SMRW)

Step 3b: Consider alternative comparison groups or other design modifications
- Average treatment effect in subset with clinical equipoise
  - Matching weights
  - Overlap weights
- Sufficient overlap achieved
- Insufficient overlap

Diagnostic step: Evaluation of balance

Mimics RCT effect
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)

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

Surah 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).

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
Results of first attempts to systematically emulate RCTs

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

Shirley V. Wang • Sushama Kattinchare 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


