UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

Gemeinsamer Bundesausschuss Innovationsausschuss

Feasibility of Benefit-Risk Evaluation of Drugs based on Non-Randomized Studies using German Registries: First Results from the NANA Project

DAGstat 25.03.2025

Paula Starke, Enes Malik Cakir, Prof. Dr. Tim Friede, Prof. Tim Mathes

Department of Medical Statistics Göttingen

The NANA project

- Goals:
  - assess the suitability of currently available registers in Germany for conducting non-randomized studies for benefit-risk evaluation of drugs
  - development of both methodology and registry quality
- Plan and conduct 10-12 studies using “state-of-the art” methodology
  - Identify question and registry
  - **Target trial + estimand**
  - **Identify confounding variables**
  - **Plan emulation**
  - **Assess feasibility**
  - Analysis

→ entire work process outlined in master protocol

→ templates for all steps

Example

- Is there an additional benefit of ozanimod compared to interferon ß-1a in adult patients with relapsing multiple sclerosis with respect to the annualized relapse rate? We emulate the existing phase-3 trials Radiance B and Sunbeam and a pragmatic trial.

Registry:

- German MS registry (DMSG)
- Documentation of about 13 000 patients with an official MS diagnosis per year

Target Trial

Objective:	**What do you want to know (in non-technical words)?** What knowledge do you need to make a better treatment decision?	
Assumed direction of effect:	**What do you expect to find out?** What is your best guess based on the existing knowledge?	
Population (key inclusion-exclusion criteria):	**Which patients are you interested in?** The results of the trial only apply to the included population, so you should describe the relevant patients as precisely as possible. List all patient characteristics in the form of criteria a patient has to fulfill to be included in the trial. Only consider characteristics that are known before the trial begins or that can be measured at the day of recruitment. Apart from this prerequisite you do not need to pay attention to the feasibility of the inclusion process as we design an ideal hypothetical trial	

Intercurrent events

Event during follow-up	might occur in study	ends follow-up	is observable in registry	end observ n i regis
Date of outcome (relapse)	yes	no	yes	no
Date of death	yes	yes	yes	ye
Patient drops out of registry	yes	yes	yes	ye
Date of add to/switch from intervention	yes	no	yes	no

Target Trial

	Pragmatic target trial of interest
Objective:	
Hypothesis:	
Population:	
Intervention:	
Comparator:	
Treatment assignment	
Outcome:	

- Are all criteria known to treating physicians at time zero? (never use info from the future!)
- Can all included patients receive BOTH therapies in practice?
- Are we interested in current users or new users only? (wash-out periods)

- What is the treatment alternative in practice? (standard of care instead of placebo)

- How long after diagnosis/hospitalisation etc is the treatment is started? (immortal time bias!)

Target Trial

	Pragmatic target trial of interest
Handling of intercurrent events:	
Follow-up period:	
Causal contrast of interest/Estimand	
Population-level summary measure of effect:	

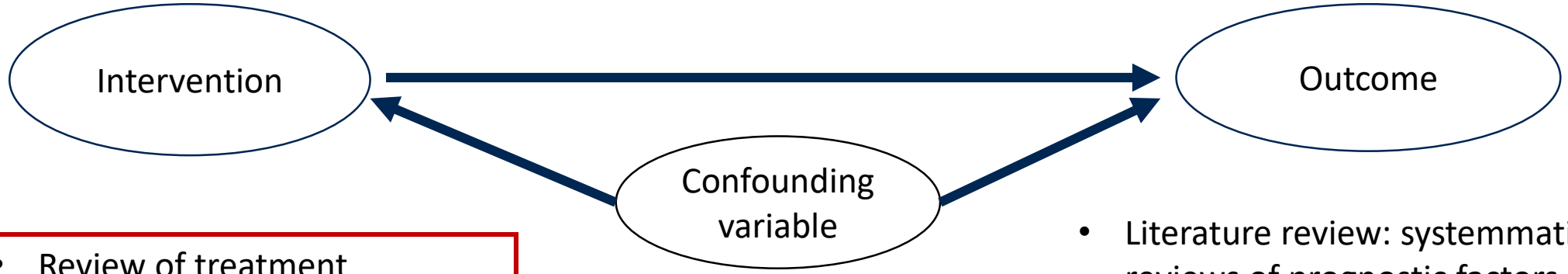
- Which events end follow-up? → choose intercurrent event strategy (see ICH E9 (R1) addendum on estimands)
- Which events lead to missing data? → choose method of imputation

- How long after start of treatment can we expect an effect to be visible?
- At which time points do patients usually show up for follow-up visits?

- We can usually only emulate the effect of initiating a treatment (similar to safety population in RCTs) instead of a real intention-to-treat effect

Adjustment set

Which criteria were really used in practice to decide which treatment a patient receives?



- Review of treatment guidelines
- Expert survey

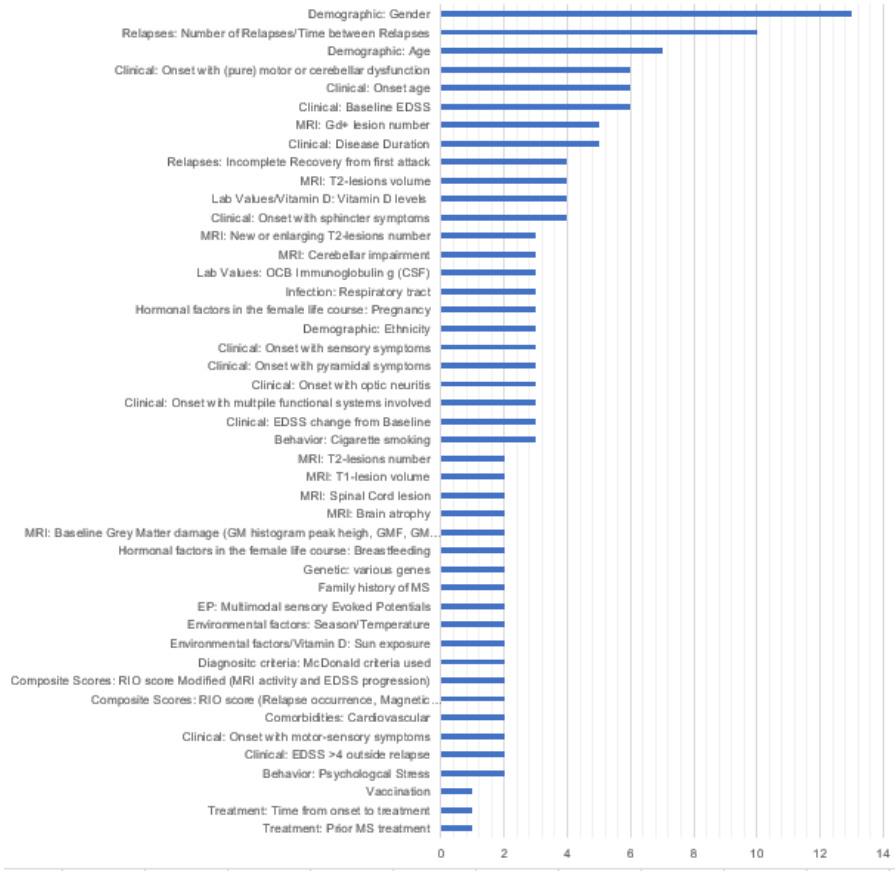
- Literature review: systemmatic reviews of prognostic factors (incl. cointerventions)
- Expert survey

Adjustment set

Expert survey

Variable	How relevant is this factor to the therapy decision if considered individually? 1… *irrelevant* 2… *little relevant* 3…*moderately relevant* 4…*rather relevant* 5…*very relevant*
relapse rate in previous year	5
pregnancy	5
number of T2 lesions	4.5
spinal or infratentorial lesions	4.5

Prognostic factors



Data we (almost) always need for successful emulations

- inclusion criteria that differentiate the population of interest from patients receiving non-relevant medications
- exact time of treatment start
- Reliably recorded outcome events with sufficient time granularity
- Time of drop-out, death and important intercurrent events (much more detailed information needed if interested in per-protocol effect or complex treatments)
- Time since onset of disease
- Important confounders

Aspects of RCTs that cannot/should not be emulated

Impossible without study context:

- Laboratory values not collected regularly in every patient
- Washout periods
- Uniform diagnostic criteria and conditions for measurements (especially imaging)
- Placebo and Blinding
- Follow-up visits at uniform times

Often not in registries:

- Basic health aspects with safety relevance (vaccinations, BMI, resting pulse, infections)
- Lab values, comorbidities, or medications that are not directly related to the investigated indication

Some insights so far

- Exact emulation of RCTs impossible → explicitely emulate hypothetical pragmatic trials
- Emulation of eligibility criteria is the first critical hurdle
- Be extra careful when defining time zero for „placebo“ (standard of care) groups
- Ask very specific questions in non-statistical language when discussing with clinicians
- Better data documentation would help
- Small sample sizes huge problem for new therapies

Further methodological goals

- prospectiv bias assessment with the new version of ROBINS-I
- simulation studies for TMLE/superlearner with small sample sizes
- Operationalize differences between RCTs and observational studies (meta-epidemiology)
- Estimands for benefit-risk assessment

Statistical analysis

Target parameter:	Average Treatment Effect as rate ratio and rate difference
Estimator and confounding adjustment:	• Doubly robust estimation using a weighted targeted maximum likelihood-estimator (wTMLE) for the rate ratio. Use propensity scores from treatment models and prognostic values from outcome models, constructed as described below. • As an alternative, treatment modelling only using inverse probability weighting.
Outcome models	**For all outcomes**, fix the following ensemble: • Main terms regression with appropriate link function and likelihood • Main terms lasso regression with appropriate link function and likelihood (via *glmnet*) • Main terms GAM with thin plate regression spline smoothers for all continuous covariates fit via restricted maximum likelihood estimation (via *mgcv*) • PolyMARS (via *polspline*) • Highly Adaptive LASSO (via *hal9001*) • Boosted trees (via *mboost::blackboost*) → superlearner
Treatment models:	Generalized linear model with main terms for all confounders.
Variance estimator:	Parametric bootstrap of targeting step