UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

Feasibility of Benefit-Risk Evaluation of Drugs based on Non-Randomized Studies using German Registries: The NANA Project

NANA Symposium Göttingen 26th May 2026

Paula Starke, Maxi Schulz, Enes Malik Cakir

(Department of Medical Statistics, University Medical Center Göttingen)

Content

- Introduction to the project
- Example study 1: Transplant Registry
- Example study 2: German Cystic Fibrosis Registry
- R packages *enfold* and *enact* for target trial emulation via TMLE and superlearning

UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

Introduction to the Project

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
  1. Identify question and registry
  2. Target trial + estimand
  3. Identification of confounding variables
  4. Assessing feasibility
  5. Analysis

  → entire work process outlined in master protocol
  → templates for all steps

Templates

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	

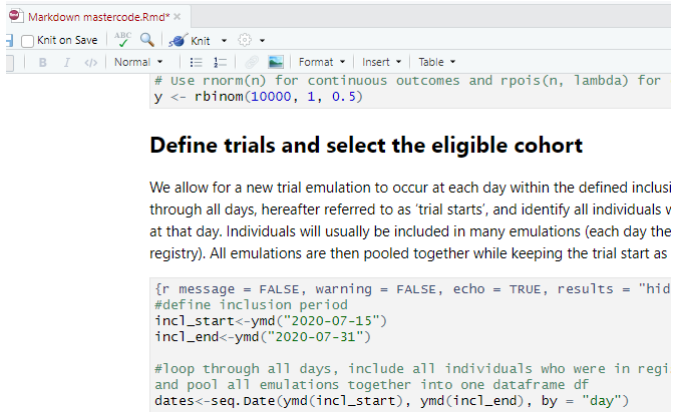
Intercurrent events

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

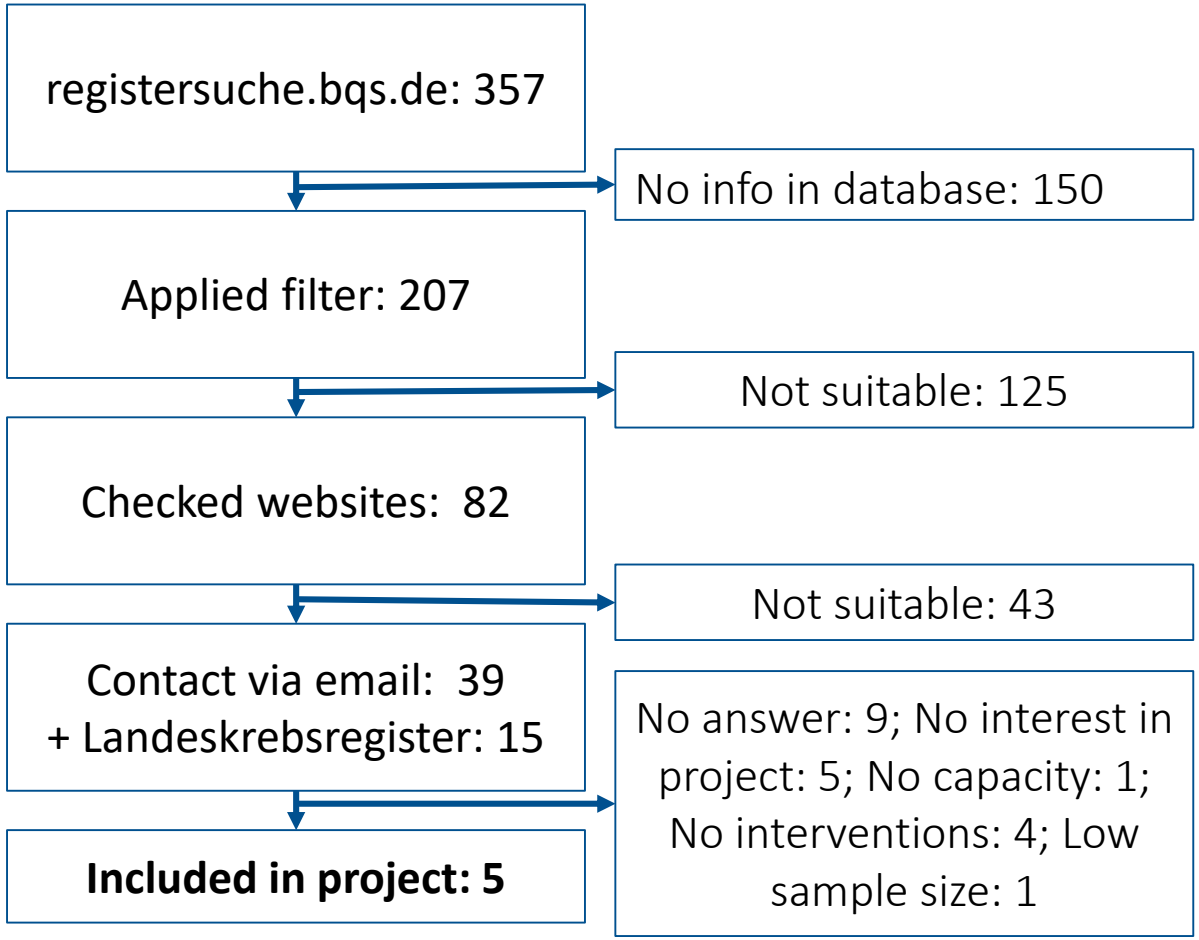
Expert survey (treatment decision factors)

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*	How strongly do you think this factor, considered individually, influences the relapse rate? *1... irrelevant* *2... little relevant* *3...moderately relevant* *4...rather relevant* *5...very relevant*
age		
sex		
brain MRI lesions at time of therapy decision		
number of T2 lesions		
volume of T2 lesions		

R-package and example scripts



Selecting registries and research questions



„MYKKE“ – Register für Kinder und Jugendliche mit Myokarditis

Deutsches Mukoviszidose-Register

DBA-Register (Diamond-Blackfan-Anämie)

MS-Register der DMSG, Bundesverband e. V.

Klinisches Krebsregister Niedersachsen

+ weitere klinische Landeskrebsregister

Selecting registries and research questions

List of currently relevant questions assembled by G-BA (federal joint commitee): 17

No registry found: 2; No interventions: 1; European registry: 1

Identified and contacted potentially suitable registries: 13

no capacity: 4; registry not suitable for this question: 2; cannot provide data: 2; no answer: 1

Included in project: 4 (+2)

Network: +2

Deutsches Myasthenie Register

Deutsches Myasthenie Register (MyaReg)

AIO

Clinical Research platform Into molecular testing, treatment and outcome of non-Small cell lung carcinoma Patients (CRISP)

TX Transplantationsregister

Paul-Ehrlich-Institut

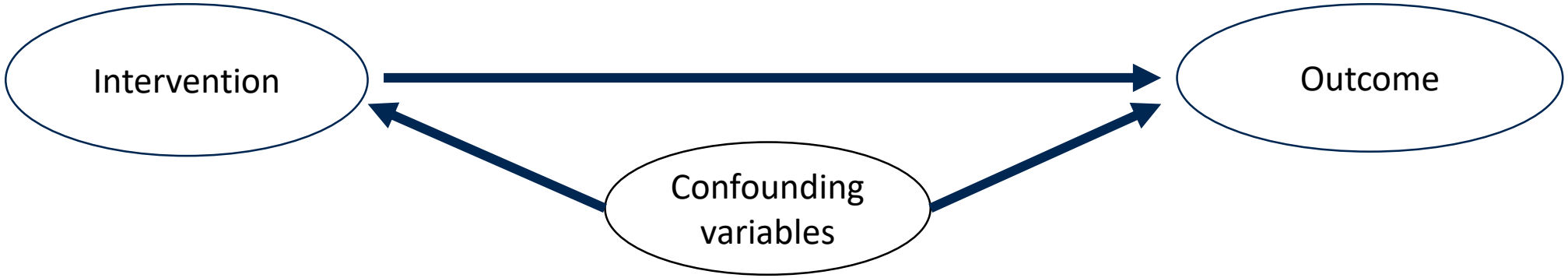
Deutsche Hämophilieregister (DHR)

Target Trials

- (Hypothetical) RCT → Emulation
  - define question in detail
  - synchronize eligibility assessment, treatment assignment and start of follow up at time 0
- Emulation of existing RCTs or new questions chosen by clinicians
- Comparison of two interventions
- Either point interventions or intention-to-treat effect → no time-varying treatments

Identification of confounding variables

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



1. Review of treatment guidelines and product information
2. Systemmatic literature review of reviews of prognostic factors
3. Summarize and categorize variable list
4. Expert survey via Excel
5. Selection of all at least little relevant variables

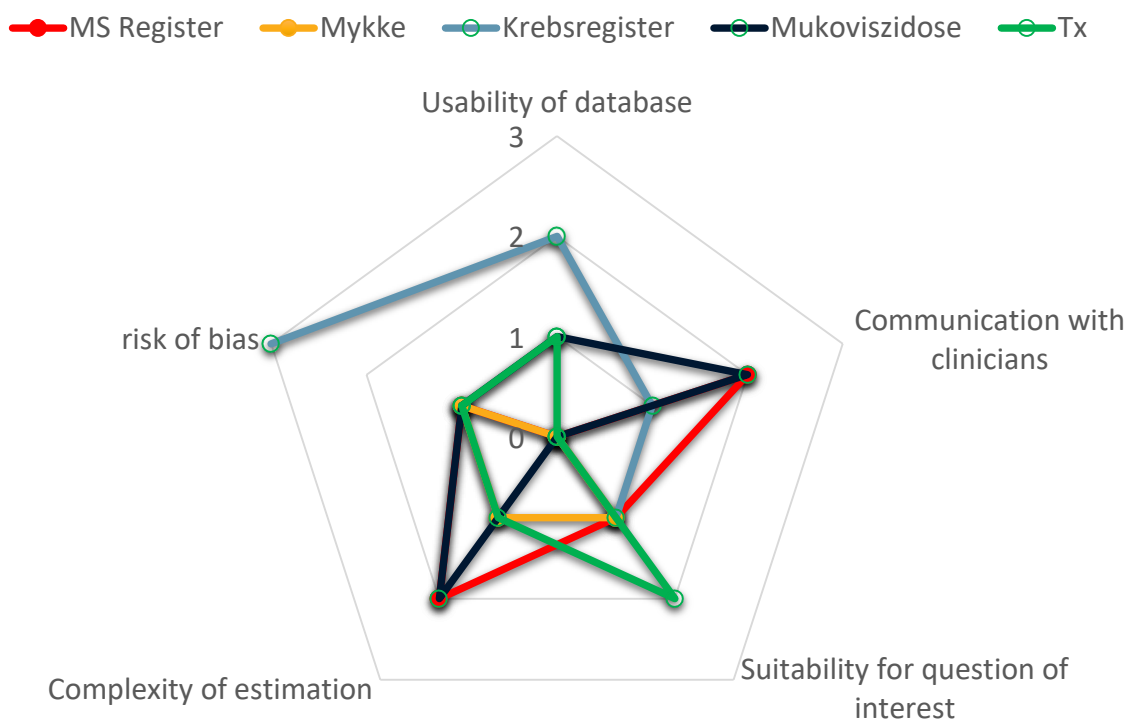
Simplified version of the approach suggested by **Pufulete et al.** (Confounders and co-interventions identified in non-randomized studies of interventions. J Clin Epidemiol. 2022;148:115-123. doi: 10.1016/j.jclinepi.2022.03.018.)

Statistical Analysis Framework

- Doubly robust estimation using a weighted targeted maximum likelihood-estimator (wTMLE) with propensity scores from treatment models and prognostic values from outcome models

Outcome models	Superlearner with a standard ensemble that can be adapted only prior to analysis
Treatment models:	Generalized linear model with main terms for all confounders
Variance estimator:	Parametric bootstrap of targeting step

Challenges (preliminary)



3 =major obstacle
2 = problematic,
1 = acceptable,
0 = very good

Challenges (preliminary)

e.g.

- Low sample size (<10 for strict inclusion criteria, <200 for pragmatic trial)
- Missing values in baseline characteristics
- Complex structure and coding of database
- Relevant endpoints not available
- Missing values in endpoints
- Short follow-up
- No dates for events during follow-up
- Very irregular follow up visits
- waiting time for data access
- No comorbidities available (relevant for eligibility criteria and confounding)

And now some example studies…

Example study 1: Transplant Registry

An emulated target trial using data of the transplant registry

Data basis transplant registry:

- Unites transplant-related medical data from various data providers (Eurotransplant, IQTIG, DSO)
- Operation since 2021 on the basis of the *Transplantation Act*
- Two phase data available: 2006-2016 (setup phase), 2017-now (regular phase)
- Available for scientific use

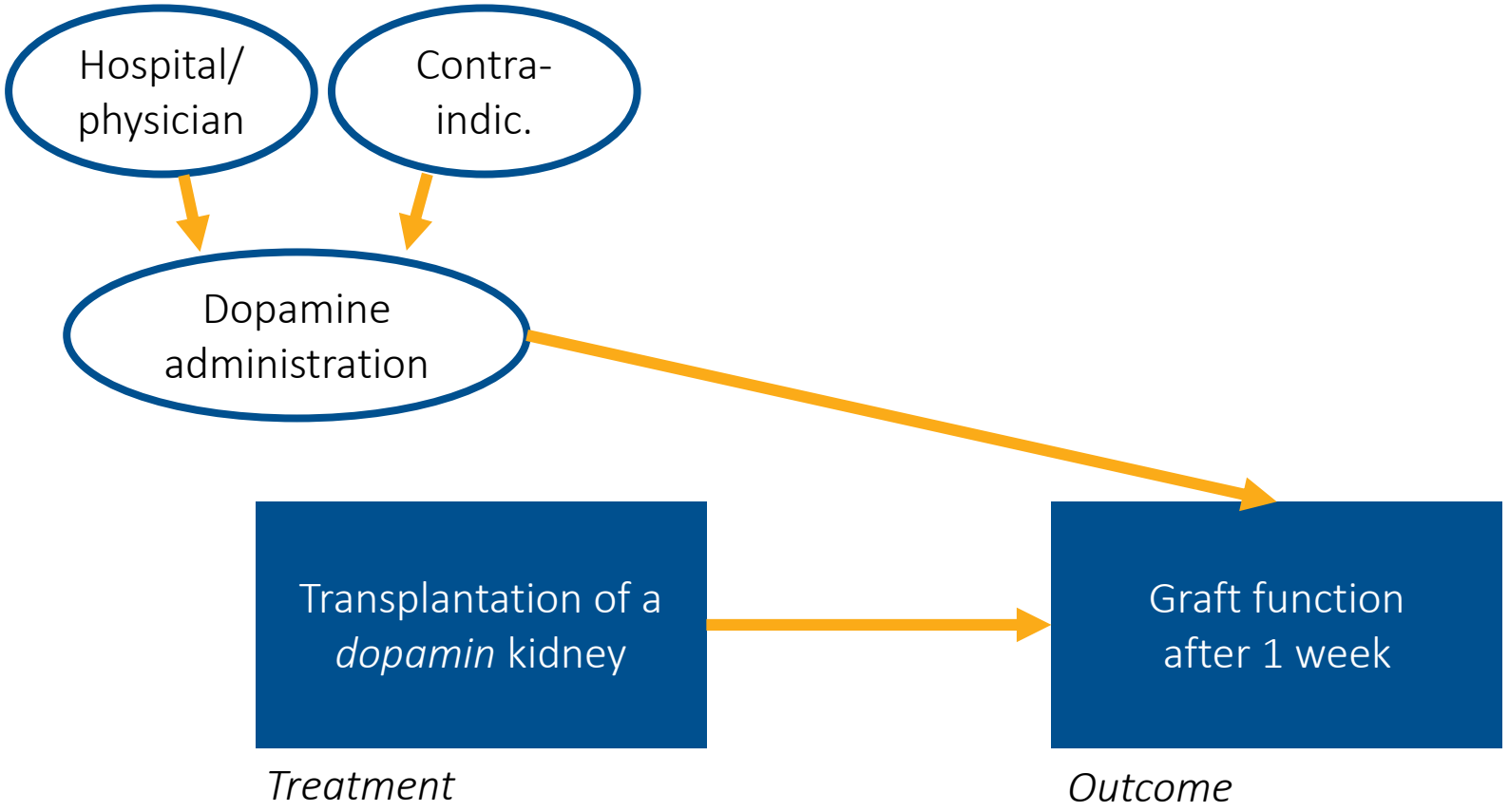
An emulated target trial using data of the transplant registry

Title: *Effects of Preoperative Administration of Dopamine to brain-dead kidney organ donors on Recipient Graft Function*

Research question:

- *Patients:* Recipients of isolated kidney transplants from brain-dead donors
- *Intervention:*
  Dopamine administration during intensive care donor management prior to organ retrieval
- *Control*:
  No dopamine administration
- *Primary endpoint:*
  Delayed graft function (defined as post-transplant dialysis within 7 days),

Confounder and time zero



- Complex identification of confounders on multiple levels (donor- and recipient-side)
- ***No Confounding*** *(transplantation as treatment)*
- ***Time zero easily identifiable,*** at time of organ transplant

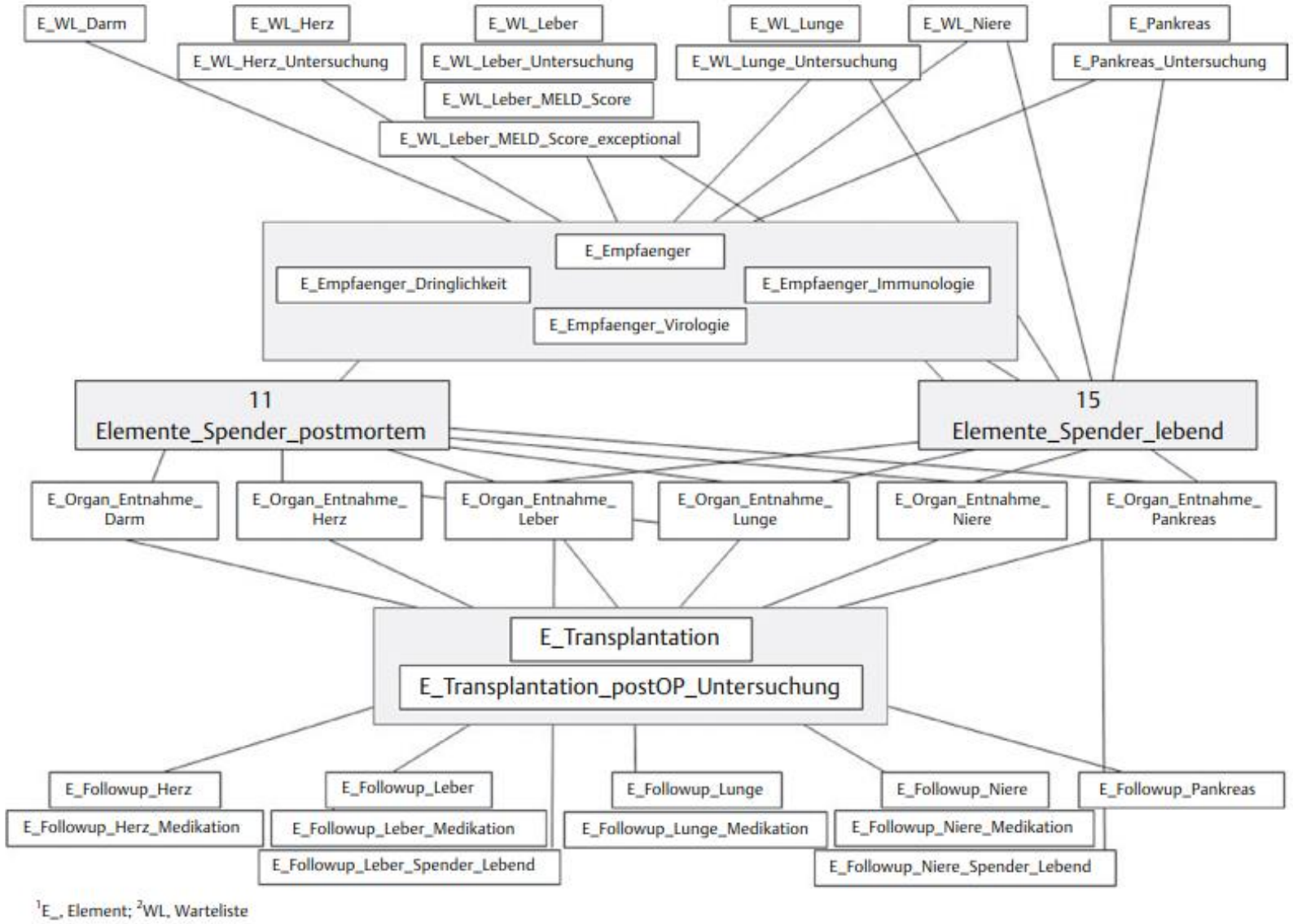


Abb. 1 Darstellung der Elemente und deren Zusammenhang im deutschen Transplantationsregister. Die aufgezeigten Verbindungen ergeben sich aus den klinischen, organisatorischen und zeitlichen Abläufen während Wartezeit, bei Transplantation und im Follow-up. Grau unterlegte Elemente sind organübergreifend angelegt, weiß unterlegte Elemente sind organspezifisch.

Gerd Otto , Klemens Budde , Christoph Bara , Jens Gottlieb, Das Deutsche Transplantationsregister – eine Analyse der Altdaten 2006–2016, in: Gesundheitswesen 2024; 86(10): Seiten 633-639.

Key challenges in data preparation

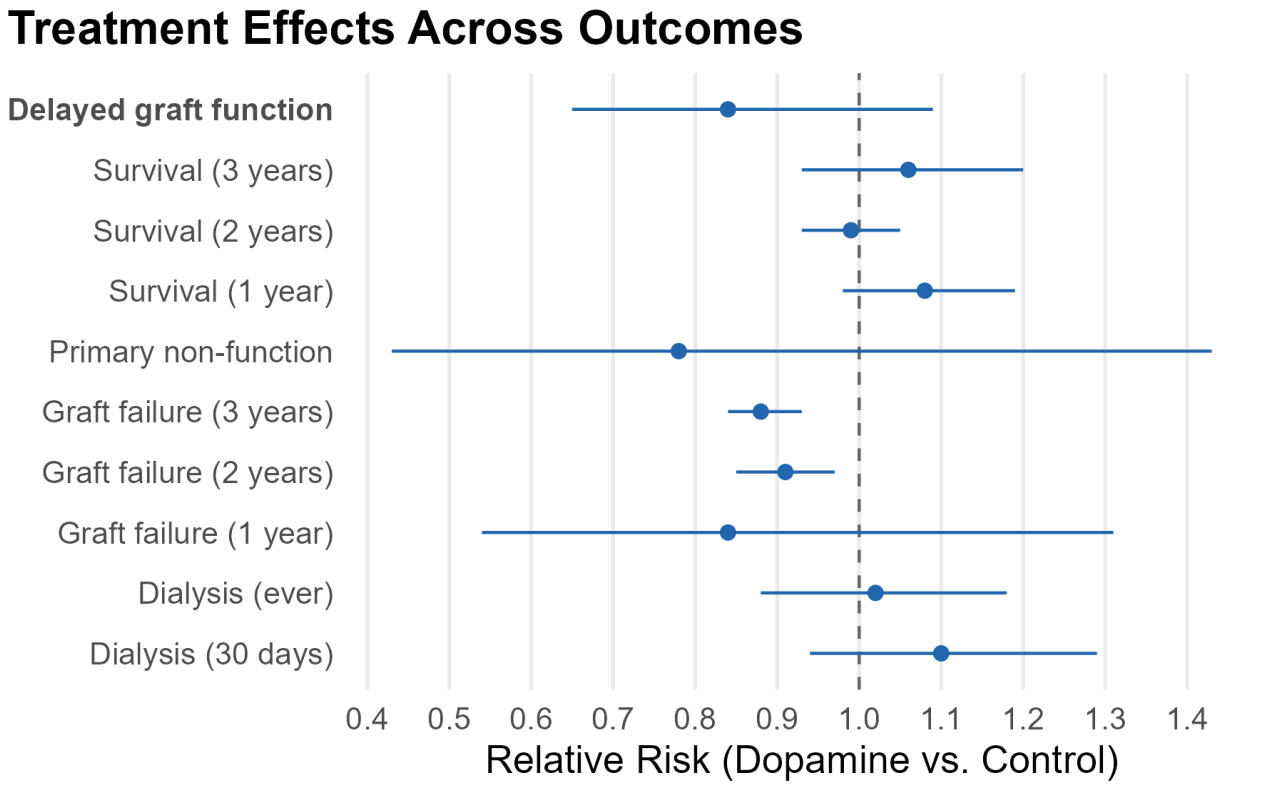
- Various data providers
  - Variety of variables for the same thing with **different levels of missingness**
  - Many variations / combinations possible
  - Knowledge of data collection necessary to assess the **reliability** of the data
- Large number of different data tables, each with a **different structure**
  - Requires comprehensive data inspection
  - Merging data often results in **missing values**, as not all records can be matched to a donor or recipient from the other table (blocks of missingness)

Difference between trial of interest and emulation

	Difference	Challenges/Problems	Solution
Objective	✓		
Hypothesis	✓		
Population	✓		
Intervention	✓		
Comparator	✓		
Treatment assignment	✓		
Outcome	✗	Primary outcome variable ▪ rarely filled, ▪ Contradictory with other variables in the dataset	New primary outcome: Composite variable with consistent behavior; inconsistencies treated as missing.
Follow-up period	✗	No timestamp available for primary outcome event (post-transplant dialysis)	Follow-up period: Adjusted to each variable's assessment window, resulting in varying follow-up durations per patient (time to hospital discharge).
Handling of intercurrent events	✗	One IC (Graft thrombosis) not identifiable	Not emulated
Causal contrast	✓		
Population-level summary	✓		

(Preliminary) Results

- Main finding: **No convincing effect of dopamine on early transplant-related endpoints.**
- Secondary finding: Association with **reduced graft failure after 2–3 years.**
- **Unexpected effect in subgroup analysis** of donors without cardiovascular comorbidity with higher survival rate in dopamine group
  - Patient survival 1 Jahr: RR 1,09 (1,04–1,14)
  - Patient survival 2 Jahre: RR 1,68 (1,23–2,28)



utcome in bold. Points = estimates; horizontal bars = 95% CIs; dashed line = reference value

Specifics of the emulated trial using data of the transplant registry

- **No confounding** identified
- **Time zero** easily identifiable
- Several **challenges in data preparation**, requires resources from both clinicians and statisticians
- Defining outcome variables becomes more difficult when
  - there is **no time stamp** for the occurrence of an event of interest or
  - there is no **clear consistent assessment window**

Example study 2: German Cystic Fibrosis Registry

An emulated target trial using data of the German Cystic Fibrosis Registry

Data Source

- Registry collecting data on patients with cystic fibrosis (CF)
- 7150 patients across 87 CF centers
- Data can be requested for scientific use

Study Background

- **CFTR modulators** are a class of drugs often used in CF treatment
- Initial EU approval of **triple therapy** with CFTR modulators elexacaftor, tezacaftor and ivacaftor in 2020
- In randomized trials and post-approval studies **clearly superior** to tezacaftor + ivacaftor and ivacaftor
- Label expansion **beyond the F508del homozygote genotype**, which dual CFTR modulators were previously restricted to

An emulated target trial using data of the German Cystic Fibrosis Registry

Title: *Effectiveness Of Elexacaftor + Tezacaftor + Ivacaftor Compared to Standard Therapy In The German Cystic Fibrosis Registry*

Research question:

- *Patients:*
  Cystic fibrosis patients with at least one F508del mutation
- *Intervention:*
  Treatment with Elexacaftor + Tezacaftor + Ivacaftor
- *Control*:
  Dynamically constructed
  F508del homozygotes: Tezacaftor + Ivacaftor
  F508del heterozygotes: Best supportive care
- *Primary endpoint:*
  Change in FEV1% predicted at week 24

MUKOVISZIDOSE e.V.
Helfen. Forschen. Heilen.

Preliminary results

- Triple therapy **superior across outcomes**
- Results for primary outcome mean triple therapy is estimated to **cause an increase in FEV1 percent predicted of about 10% more than the comparator after 24 weeks** (estimates compatible with increases in the range 5% to 13.7%)
- Triple therapy also lead to fewer exacerbations and an increase in BMI

Outcome name	Estimated average treatment effect with confidence interval
Change in FEV1% predicted	9.38 [5.06, 13.71]
Difference in FEV1/FVC ratio	0.035 [0.029, 0.041]
Number of exacerbations	-0.75 [-0.98, -0.51]
BMI	0.64 [0.54, 0.73]

Challenge: time zero

- In target trial emulations, we need to align three time points **for each intervention**:
  - Start of follow-up in the registry
  - Time at which eligibility criteria are met
  - Assignment of the intervention
- Not aligning all three time points → bias
- **Not possible for all research questions or all data sources**

In the cystic fibrosis registry:

- Decision in favor of best supportive care can only be made during center visits
- CFTR modulator start dates **not necessarily on center visits** → not synchronized

(Imperfect) solution: match CFTR modulator starts to a recent clinic visit, if possible

Challenge: software implementation

Complex intervention:

- Comparison group constructed **dynamically**:
  - **Tezacaftor + ivacaftor** for F508del homozygotes
  - **Best supportive care** for F508del heterozygotes

→ Requires different estimation machinery **not provided** by many packages for doubly robust estimation

Protocol complexity:

- *R* packages for statistics in causal inference typically not made with complex study protocols in mind
  - Often no way to separate inverse probability weight estimation from inference (necessary for checkpointing)
  - Emulations with many outcomes create many intermediate objects that are tedious to track

UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

R packages *enfold* and *enact* for target trial emulations via TMLE and superlearning

Emulating target trials in *R*

Remember the cystic fibrosis target trial required emulating a dynamic intervention. Trying to emulate such a target trial in *R*, we will find:

- **Most popular packages** assume an overly simplistic problem structure and fail
- **More flexible packages exist**, but they subsume all analysis steps into one big function

enfold and ***enact*** seek to address these problems by emulating complex trials over many small steps.

- *enfold* is a modern, efficient and powerful package for **model ensembling** and **superlearning** tailored towards causal inference
- *enact* is a **target trial composition package** with support for arbitrary interventions and reporting helpers

Why use *enfold*?

enfold is **feature-rich**:

- Beyond fitting standard learners, has native support for **pipelines**, **grids** and **hyperparameter optimization**
- Has an innovative **list learner** concept that allows providing an arbitrary number of predictions for a single learner
  - Useful for algorithms that are fit across regularization paths

enfold is **efficient**:

- Support for file-backed data via *bigstatsr* (matrices) and *arrow* (data frames)
- Support for parallelization via futures

enfold is **flexible**:

- Easy creation of custom learner templates

Why use *enact*?

Complex target trials often need to be emulated with a lot of custom code called across many outcomes, and requires figuring out how to interface your results with other packages. *enact*:

- Creates a special type of object that **evolves** as parts of the emulation are implemented
- Explicitly **separates** tasks during analysis which belong to different parts of a study protocol
- Allows for very convenient wrappers for **analysis** and **reporting**

Example: Most packages estimate propensity scores and causal estimates in the same step. In *enact*, they are estimated separately. **All intermediate results are stored in the same list** and can be accessed from there.

Example: creating learners in *enfold*

Let us first define learners and a metalearner. We create a GLM and a random forest using pre-defined templates. We also make a superlearner.

```
my_ranger <- lrn_ranger(name = "Ranger", num.trees = 1000)
my_glm <- lrn_glm(name = “GLM”, family = binomial())

my_learners <- list(my_ranger, my_glm)
my_sl <- mtl_superlearner(name = "SL", loss_fun = loss_logistic())
```

Example study via *enact*

If we specify which columns are confounders, treatments and outcomes and add labels, *enact* can call *gtsummary* to create a table 1 according to your instructions:

```
my_tte <- initiate_study(
  data = tte_data,
  confounders = c("L1", "L2"),
  confounder_labels = c("Confounder 1", "Confounder 2")
) |>
  add_treatment("A", label = "Treatment") |>
  add_outcome("Y", label = "Outcome", censoring = NULL) |>
  create_table_one()
```

Table 1 created via *enact*

Tables created by *enact* use task information to structure themselves. The package uses *gtsummary* under the hood.

Table 1. Baseline characteristics

	Treatment		
Characteristic	**Overall** N = 100[1]	**Treatment: No** N = 60[1]	**Treatment: Yes** N = 40[1]
Confounder 1	0.03 (1.04)	-0.07 (1.00)	0.19 (1.09)
Confounder 2	-0.09 (0.90)	-0.21 (0.88)	0.09 (0.92)

[1] Mean (SD)

Example study via *enact*

If we have balance checks at a checkpoint, we can define interventions, add models, and fit interventions without yet fitting outcomes or estimating any target parameter.

```
my_tte <- my_tte |>
  define_interventions(
    static_intervention(0, "Control"),
    static_intervention(1, "All treated")
  ) |>
  add_models(
    treatments(learners = my_learners, metalearner = my_sl),
    outcomes(learners = my_learners, metalearner = my_sl)
  ) |>
  fit_interventions() |>
  check_balance()

summary(my_tte$balance_checks)
```

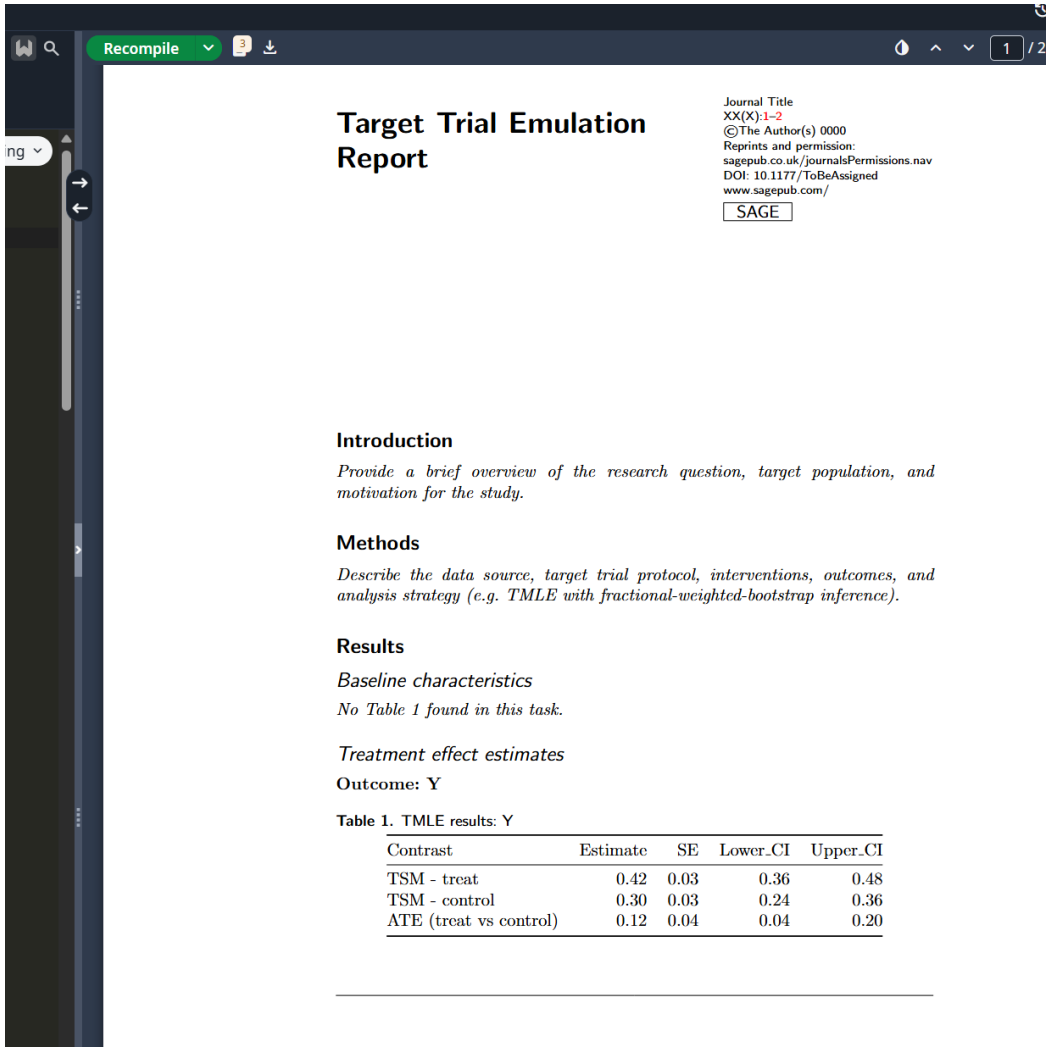
Example study via *enact*

Now we only need to fit outcomes and do targeting. At task end, we can:

- Call strip() to remove all sensitive information from a task, retaining only results that are safe to send to colleagues
- Call report() to send results to a .docx or .tex file from which you can prepare a manuscript

```
my_tte <- my_tte |>
  fit_outcomes() |>
  do_tmle(which = all_outcomes(), fluctuation_family = binomial()) |>
  strip() |>
  report(template = rticles::sage_article())
```

Report created by *enact*



Target Trial Emulation Report

Journal Title
XX(X):1–2

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Introduction

Provide a brief overview of the research question, target population, and motivation for the study.

Methods

Describe the data source, target trial protocol, interventions, outcomes, and analysis strategy (e.g. TMLE with fractional-weighted-bootstrap inference).

Results

Baseline characteristics

No Table 1 found in this task.

Treatment effect estimates

Outcome: Y

Table 1. TMLE results: Y

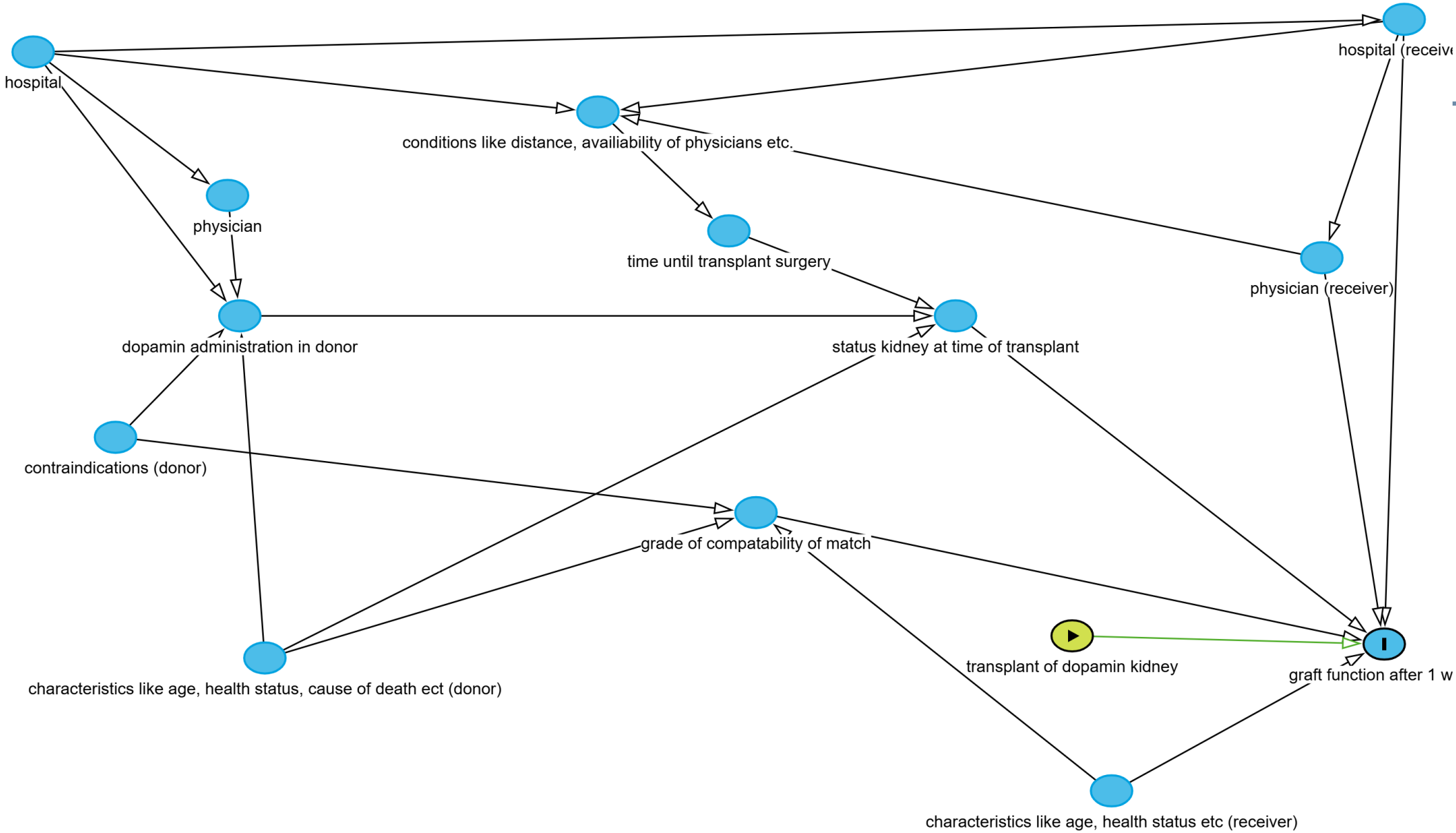
Contrast	Estimate	SE	Lower_CI	Upper_CI
TSM - treat	0.42	0.03	0.36	0.48
TSM - control	0.30	0.03	0.24	0.36
ATE (treat vs control)	0.12	0.04	0.04	0.20

Summary

Rather than doing a custom implementation or using big opaque functions, *enfold* and *enact* let you:

- solve tasks **in the order they appear in a study protocol**,
- specify arbitrarily simple or complex interventions,
- **bundle all results** automatically in one object,
- use strip() and report() to send results to a file for manuscript writing.

Additional material


hospital
hospital (receiv
conditions like distance, availiability of physicians etc.
physician
time until transplant surgery
physician (receiver)
dopamin administration in donor
status kidney at time of transplant
contraindications (donor)
grade of compatability of match
transplant of dopamin kidney
graft function after 1 w
characteristics like age, health status, cause of death ect (donor)
characteristics like age, health status etc (receiver)