UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

AI/ML IN CLINICAL RESEARCH

Artificial Intelligence, Causality and Personalized Medicine (AICPM 2022)

Tim Friede
Institut für Medizinische Statistik
Universitätsmedizin Göttingen

APPLICATIONS OF AI IN CARDIOVASCULAR MEDICINE

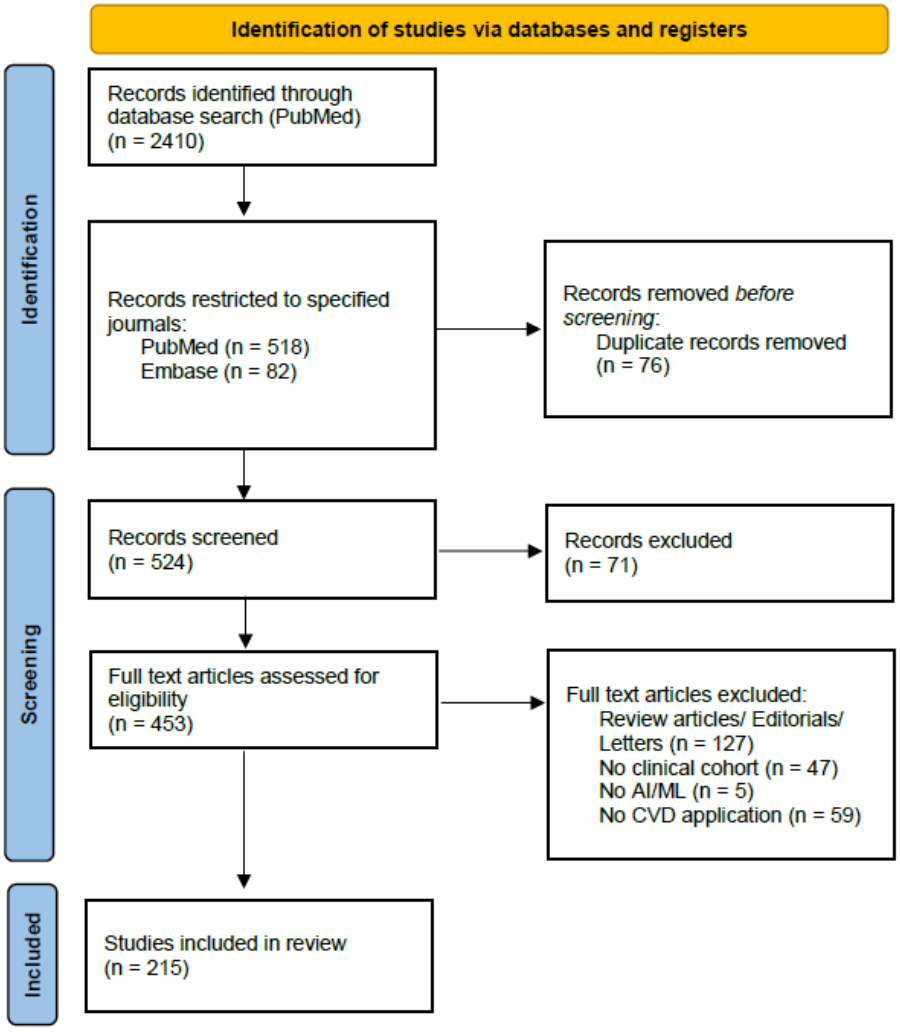
- DZHK (German Center for Cardiovascular Research)
- Project group AI / ML

ESC European Society of Cardiology
European Heart Journal Digital Health
doi:10.1093/ehjdh/ztab054
REVIEW

Applications of artificial intelligence/machine learning approaches in cardiovascular medicine: a systematic review with recommendations

Sarah Friedrich [1]*, Stefan Groß [2,3], Inke R. König[4,5], Sandy Engelhardt[6,7,8], Martin Bahls[2,3], Judith Heinz[1], Cynthia Huber[1], Lars Kaderali[3,9], Marcus Kelm [10,11,12,13], Andreas Leha[1,14], Jasmin Rühl[1], Jens Schaller[10,13], Clemens Scherer [15,16], Marcus Vollmer[3,9], Tim Seidler[14,17], and Tim Friede [1,14]

SYSTEMATIC REVIEW



- **Objectives**: (a) describe current state of AI/ML applications in CV medicine, and (b) provide recommendations for future applications
- **Search strategy**: Pubmed and EMBASE searched for publications using AI/ML approaches in CV medicine from 2000 onwards (last search March 2020)

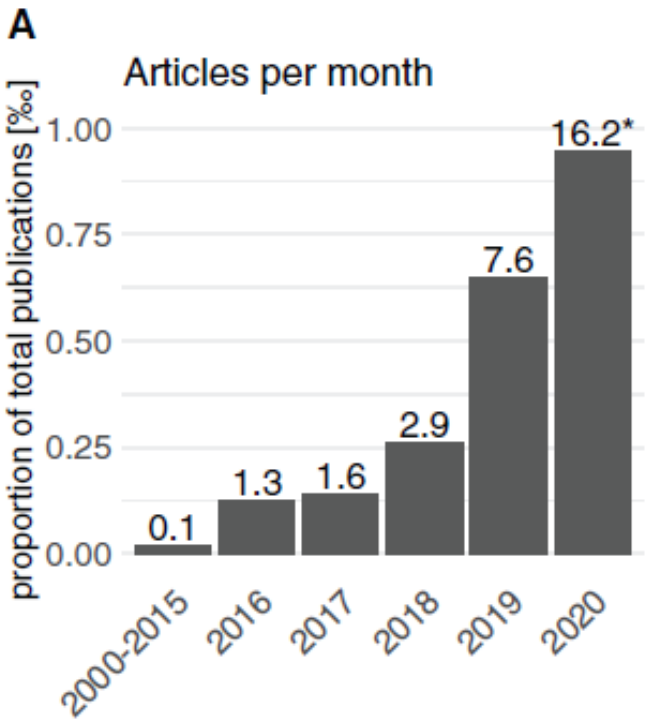
SYSTEMATIC REVIEW

- Study characteristics and temporal trend in publications

Table I **Study characteristics**

Variable	Level	Total
Subjects	Median (IQR)	1083.0 (213.5–10 757.0)
Subject categories	<100	31 (14.4)
	100–1000	73 (34.0)
	1000–10 000	53 (24.7)
	10 000–100 000	45 (20.9)
	100 000–1 000 000	11 (5.1)
	>1 000 000	2 (0.9)
Design	Prospective cohort study	48 (22.3)
	Retrospective cohort study	138 (64.2)
	Case-control study	20 (9.3)
	RCT	9 (4.2)
Outcome	Binary	153 (71.2)
	Categorical	13 (6.0)
	Continuous	27 (12.6)
	Time to event	22 (10.2)
Secondary data	No	64 (29.8)
	Yes	151 (70.2)

Values are *n* (%) unless otherwise stated.



SYSTEMATIC REVIEW

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

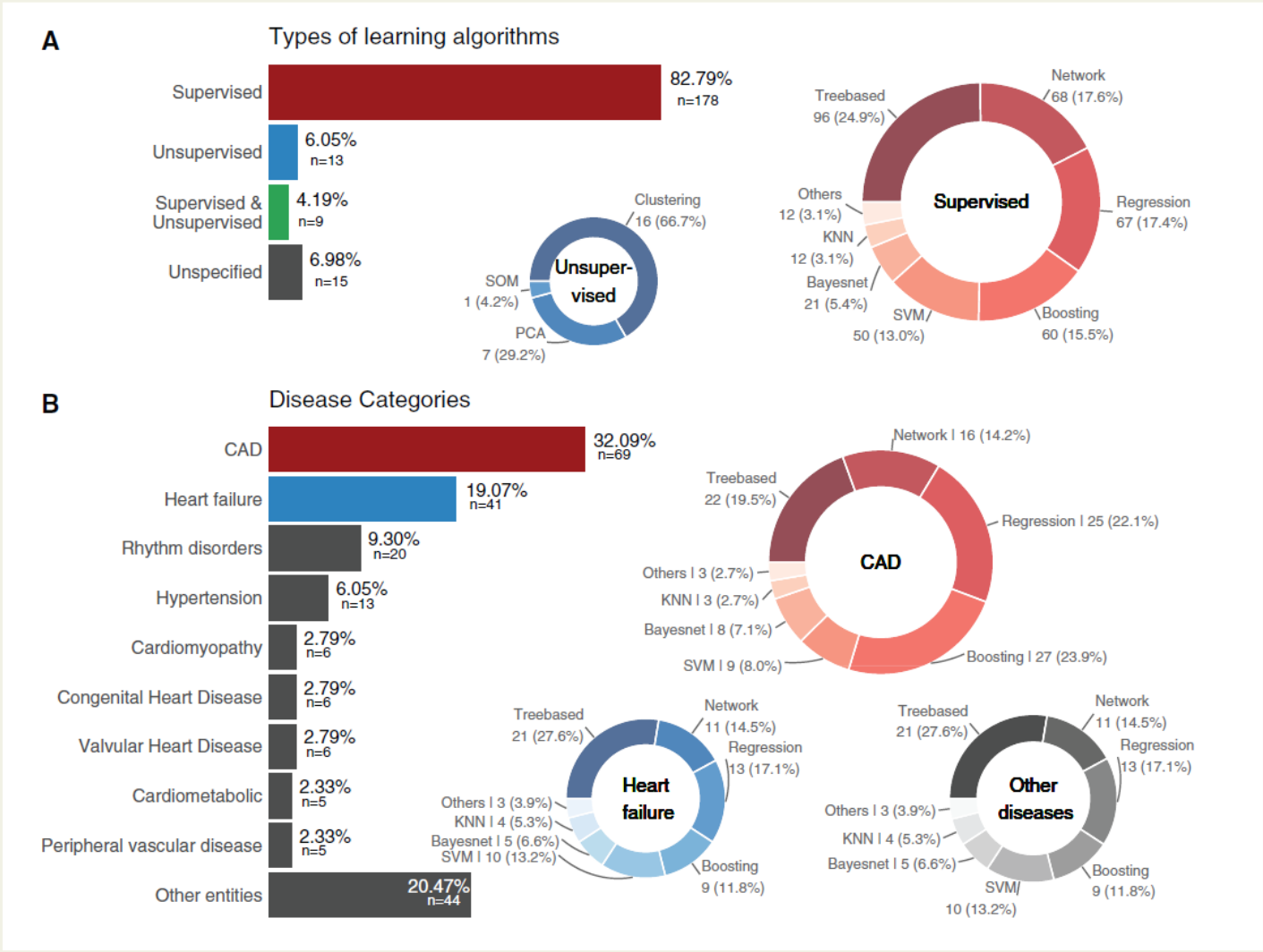


Figure 2 Overview of the methods and disease areas presented in the articles. Panel (*A*) shows the types of artificial intelligence/machine learning algorithms applied. Panel (*B*) displays the distribution of disease areas as well as which supervised methods are most commonly applied in which disease area.

SYSTEMATIC REVIEW

- **Types of data used**

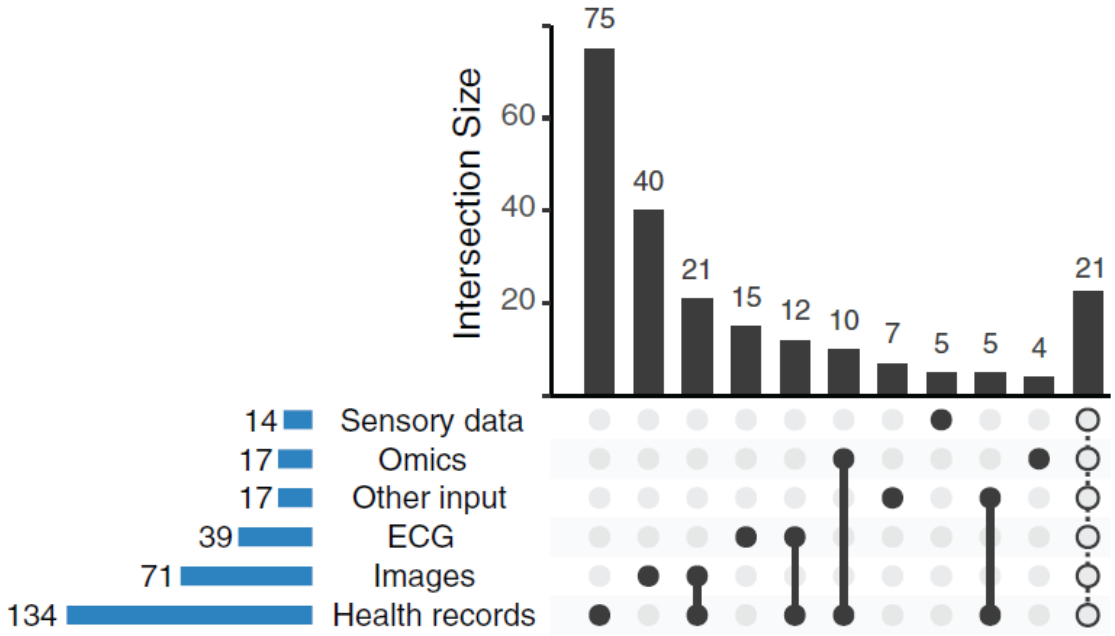
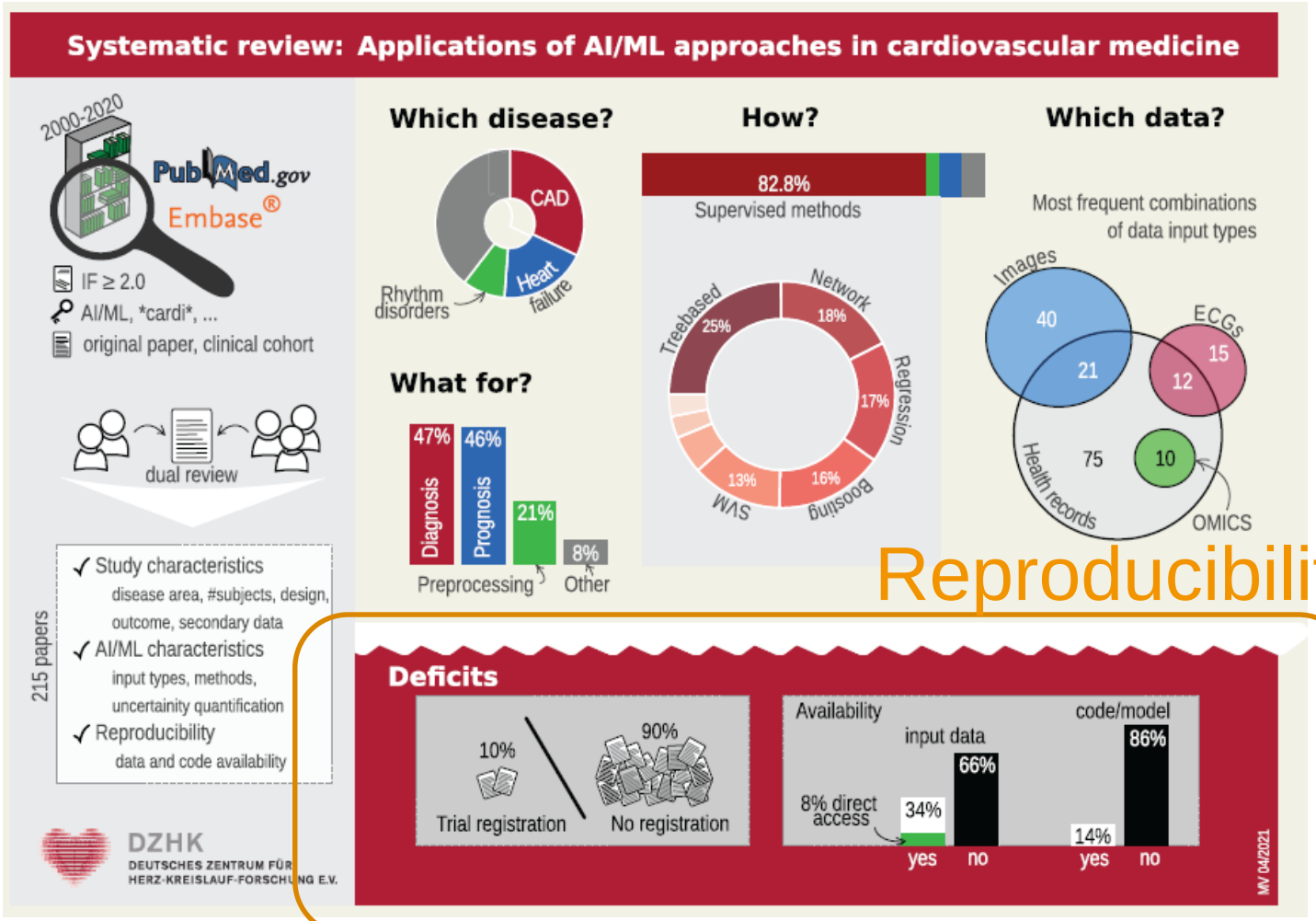


Figure 3 Input type used for the artificial intelligence/machine learning algorithms. Displayed are the absolute number a respective input type was used (lower left bars) and the most common combinations of input types (upper bar plot). The last bar summarizes all other combinations that occurred less than four times.

SYSTEMATIC REVIEW

- **Summary**: graphical abstract



RECOMMENDATIONS

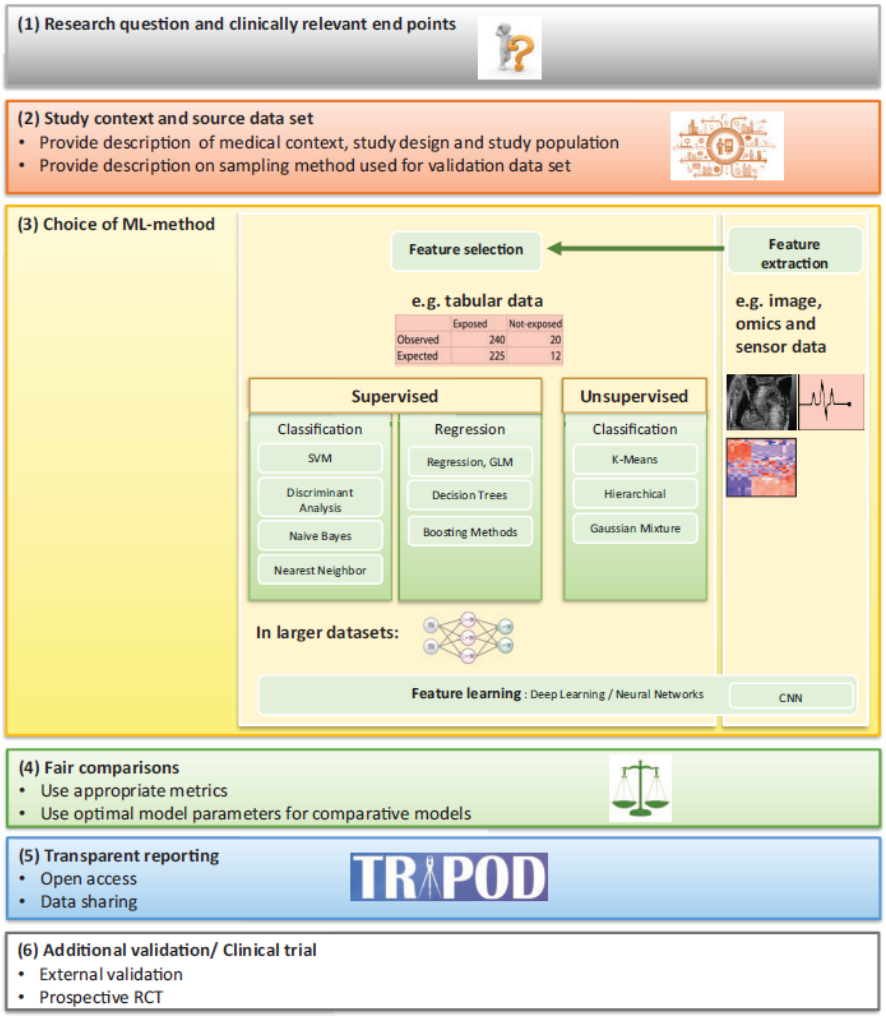


Figure 5 Recommended steps to be taken into account when using artificial intelligence/machine learning methods in cardiovascular research. Feature selection (selecting the most relevant subset of features, e.g. a biomarker, age or sex of a patient or image information), feature extraction (finding a minimalistic representation of a larger data set, e.g. an image), and feature learning (the algorithm chooses/learns relevant features from the data).

DAGSTAT WHITE PAPER

- **DAGStat (German Consortium of Statistics)**
  - Association of 13 professional and learned societies and the Destatis (Federal Statistical Office in Germany)
  - Homepage https://www.dagstat.de/

Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-021-00455-6

REGULAR ARTICLE

Is there a role for statistics in artificial intelligence?

Sarah Friedrich, et al. *[full author details at the end of the article]*

Received: 13 September 2020 / Revised: 6 July 2021 / Accepted: 7 July 2021

THE STATISTICAL PERSPECTIVE

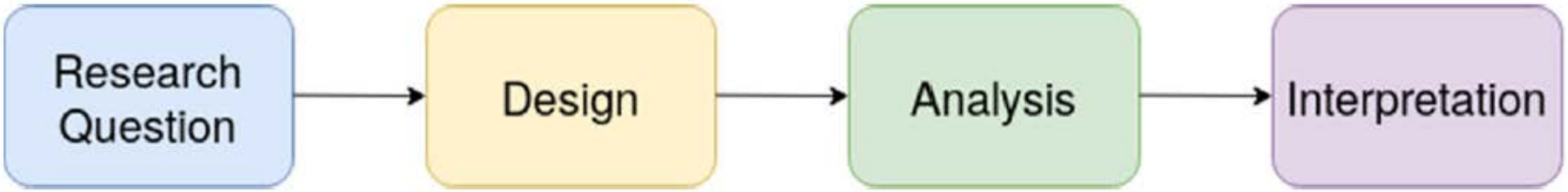


Fig. 1 Flow chart of study planning, design, analysis and interpretation

STATISTICAL THINKING TO IMPROVE ARTIFICIAL INTELLIGENCE METHODS AND APPLICATIONS

- **Design**: bias reduction; validation; representativity; selection of variables
- **Assessment of data quality**: standards for the quality of diagnostic tests and audits; dealing with missing values
- **Differentiation between causality and associations**: consideration of covariate effects; answering causal questions; simulation of interventions
- **Assessment of certainty or uncertainty in results**: Increasing interpretability; mathematical validity proofs or theoretical properties in certain AI contexts; providing stochastic simulation designs; accurate analysis of the quality criteria of algorithms in the AI context

STATISTICAL APPROACHES FOR STUDY DESIGN AND VALIDATION

- **Experiment - observational study - convenience sample**
  - AI utilizes often convenience samples (e.g. routine data) since large data sets and accessible at low costs
- **Large data sets not necessarily representative** of a (target) population
- **Example:** Apple Heart Study (Perez et al (2019) NEJM)
  - Objective: To assess Apple Smartwatch‘s ability to identify atrial fibrillation
  - Large-scale assessment including more than 400,000 participants
  - Caveat: Average age of participants 41 years, but AF most prevalent in older patients (>65 years of age)

STATISTICAL STUDY DESIGN

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- **Representativity**
  - Naive expectation that sufficiently large data automatically lead to representativity (Meng 2018; Meng and Xie 2014)
  - Careful planning required
- **Bias** (such as selection, attribution, performance, and detection bias)
  - Statistical methods and principles for minimizing bias, e.g. stratification (Simpson's paradox)
- **Sample size planning**

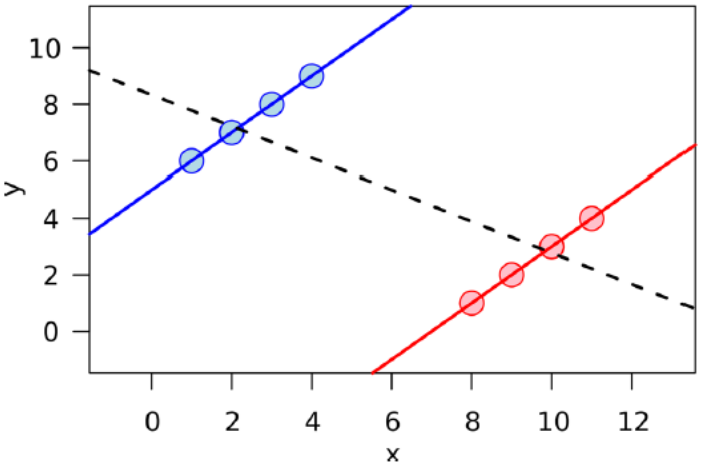


Fig. 2 Simpson's paradox for continuous data: a positive trend is visible for both groups individually (red and blue), but a negative trend (dashed line) appears when the data are pooled across groups (Wikipedia 2020) (color figure online)

STATISTICAL APPROACHES FOR VALIDATION

- **Internal vs. external validation**
- **Bench marking** (computer science) vs. **simulations** (statistics) (see also https://arxiv.org/abs/2208.01457)
  - Experience with structuring, reporting and interpreting simulation studies in statistics
- **Sample sizes**
  - Requirements depending on dimensionality, sparsity, non-linearity, …
  - Sample size planning common task in clinical trials, but not routinely performed in AI / ML applications
- **Fast development cycles with AI technologies** (often faster than validation studies)

ASSESSMENT OF DATA QUALITY

- **‘Data is the new oil of the global economy.’**
  - Not really, crude oil needs refining and is limited
  - Still highlights the importance of data for the economy
- **‘Garbage in, garbage out.’**
  - More data ≠ more information (e.g. random-effects meta-analysis; Jackson and Turner, 2017)

ASSESSMENT OF DATA QUALITY

- AI often utilizes **data lakes** (vast amount, collected for different purpose, convenience samples, …)
- 'Extract, Transform, Load' (**ETL process**)
- AI need to be trained and evaluated on **'fit for purpose' data** (includes relevance, completeness, availability, timeliness, meta-information, documentation and context-dependent expertise) (Duke-Margolis, 2018)

Fig. 3 Data relevancy and quality are equivalent components of a fit-for-purpose real-world data set. Figure according to Duke-Margolis (2018)

ASSESSMENT OF DATA QUALITY

- In **official statistics**: similar concepts of data quality exist
- For instance, **dimensions of data quality**, including relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity (European Statistical System, 2019)
- **Exploratory data analysis** (including visualization)
  - Pre-processing to detect anomalies or to define ranges of typical values in order to correct input or measurement errors and to determine standard values
  - Applying complex black-box methods without pre-processing dangerous

CAUSALITY AND ASSOCIATION

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

- **AI / ML excellent at discovering associations**
- Important to acknowledge in the interpretation of results: **Association not necessarily due to causal relationship**
- **Lessons to be learned from other fields**
  - Clinical epidemiology: Bradford Hill criteria
  - Statistics: Counterfactual framework by Rubin (1974)

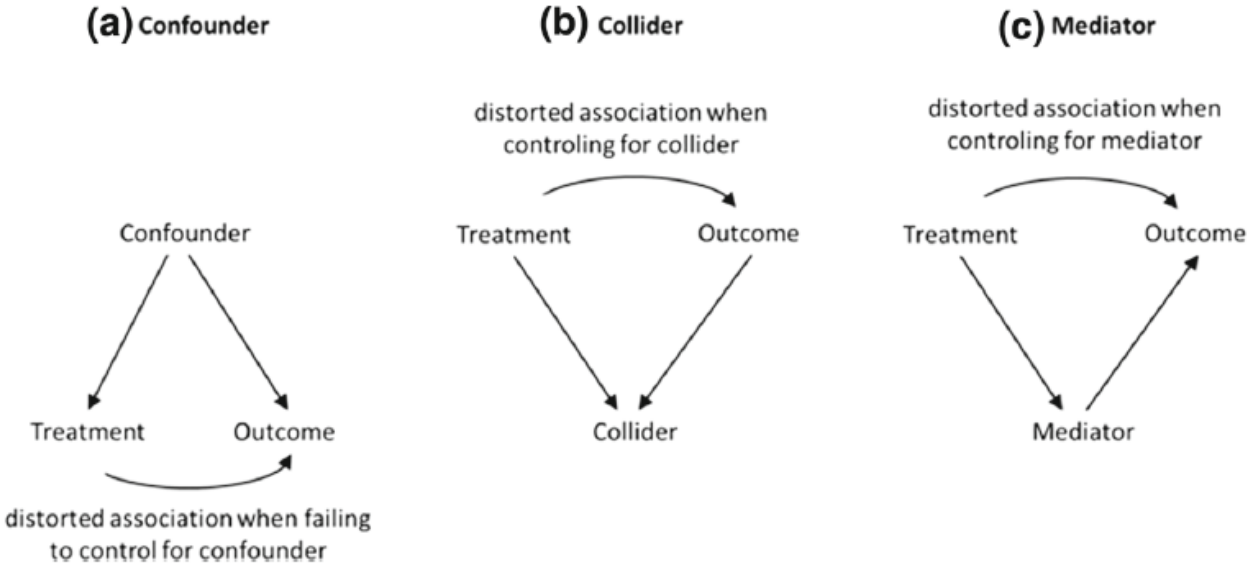


Fig. 4 Covariate effects in observational data, according to Catalogue of bias collaboration (2019)

EVALUATING UNCERTAINTY

- **Uncertainty quantification is often neglected in AI applications**
- With large data sets **sampling variation** might be small
- However: **Model uncertainty** might remain
- Derivation / computation of measures of uncertainty such as standard errors or prediction intervals can be tricky with complex analysis methods
- More recently, application of **resampling techniques** such as bootstrapping or jack knifing (theoretical properties not always established)
- Alternative approaches embed algorithmic methods in statistical models

PERSONALIZED MEDICINE

- **Efficacy, safety and consequently benefit-risk might vary across patient population**
  - Stratification of patient populations
  - Drive towards targeted treatments
- **Implications for clinical research**
  - Identification of patient subgroups
  - Enrichment of clinical study populations

IDENTIFYING PATIENT SUBGROUPS

- **Usually requires more data than one randomised controlled trial** (Huber et al, 2019)
- **ML methods in meta-analytic framework to identify patient subgroups**

Advances in Data Analysis and Classification (2022) 16:797–815
https://doi.org/10.1007/s11634-021-00458-3

REGULAR ARTICLE

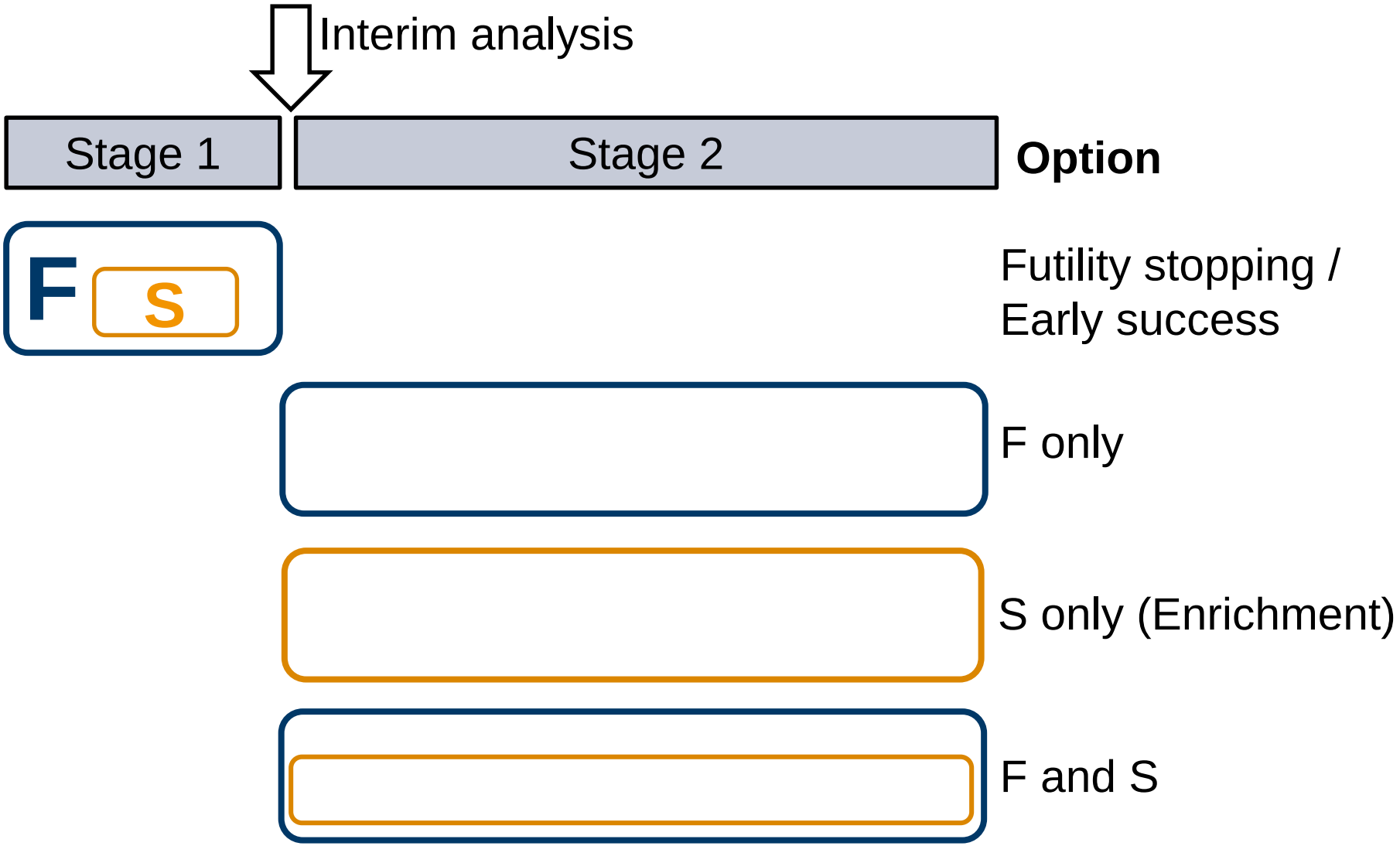
Check for updates

Subgroup identification in individual participant data meta-analysis using model-based recursive partitioning

Cynthia Huber[1] · Norbert Benda[1,2] · Tim Friede[1]

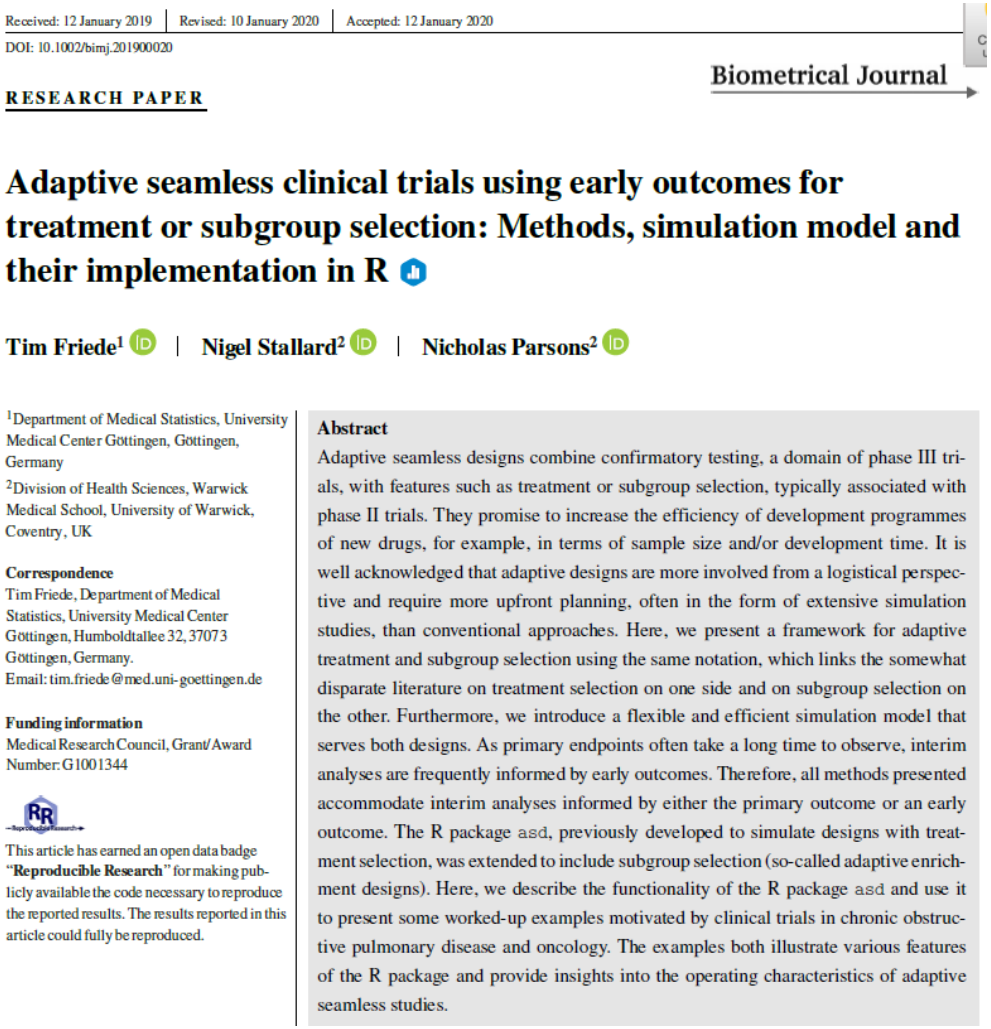
Received: 21 September 2020 / Revised: 25 June 2021 / Accepted: 22 July 2021 /
Published online: 14 August 2021

ADAPTIVE ENRICHMENT DESIGN



Friede et al. (2012) Stat Med

SIMULATING CLINICAL TRIALS



Received: 12 January 2019 | Revised: 10 January 2020 | Accepted: 12 January 2020
DOI: 10.1002/bimj.201900020
Biometrical Journal
RESEARCH PAPER

Adaptive seamless clinical trials using early outcomes for treatment or subgroup selection: Methods, simulation model and their implementation in R

Tim Friede[1] | Nigel Stallard[2] | Nicholas Parsons[2]

[1]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[2]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK

Correspondence
Tim Friede, Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany.
Email: tim.friede@med.uni-goettingen.de

Funding information
Medical Research Council, Grant/Award Number: G1001344

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Adaptive seamless designs combine confirmatory testing, a domain of phase III trials, with features such as treatment or subgroup selection, typically associated with phase II trials. They promise to increase the efficiency of development programmes of new drugs, for example, in terms of sample size and/or development time. It is well acknowledged that adaptive designs are more involved from a logistical perspective and require more upfront planning, often in the form of extensive simulation studies, than conventional approaches. Here, we present a framework for adaptive treatment and subgroup selection using the same notation, which links the somewhat disparate literature on treatment selection on one side and on subgroup selection on the other. Furthermore, we introduce a flexible and efficient simulation model that serves both designs. As primary endpoints often take a long time to observe, interim analyses are frequently informed by early outcomes. Therefore, all methods presented accommodate interim analyses informed by either the primary outcome or an early outcome. The R package `asd`, previously developed to simulate designs with treatment selection, was extended to include subgroup selection (so-called adaptive enrichment designs). Here, we describe the functionality of the R package `asd` and use it to present some worked-up examples motivated by clinical trials in chronic obstructive pulmonary disease and oncology. The examples both illustrate various features of the R package and provide insights into the operating characteristics of adaptive seamless studies.

- **Simulation model**
  - Adaptations informed by early outcomes
  - Aggregate rather than individual participant data
  - Multivariate normal model
- **R package asd** by Nick Parsons (Warwick) available from CRAN

Friede et al (2020) Biom J

IMPROVING CLINICAL TRIAL DESIGNS THROUGH BAYESIAN OPTIMIZATION

- **Optimizing adaptive seamless designs** requires simulations
- Simulations with **grid searches** can be **expensive**
- **Idea:** application of efficient optimization technique within the CSE framework
- **Bayesian optimization (BO)**
  - also known as model-based optimization (MBO)
  - to optimize expensive (time-consuming) black-box functions by using a regression as a surrogate to guide the search
  - Implementation: R-package mlrMBO (Bischl et al., 2017)
- **Reference**: Richter et al (2022) Biom J

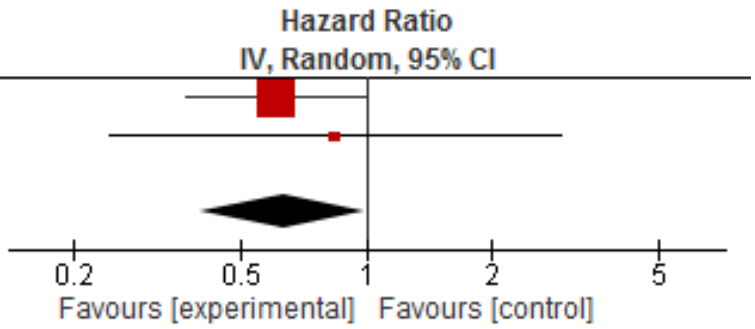
CROSS DESIGN SYNTHESIS

UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

- **Creutzfeldt-Jakob disease (CJD)**
  - prevalence of 1–9 cases per 1,000,000 people
  - qualifies as **rare disease** (EU: less than 5 in 10,000)
- **Varges et al (2017)** investigated Doxycycline in early CJD
  - double-blinded randomized phase II trial (n=12)
  - observational study (n=88) (Cox regression stratified by terciles of the propensity scores)
  - survival time as primary outcome

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Observational study	-0.49948	0.2493	86.5%	0.61 [0.37, 0.99]
RCT	-0.17344	0.6312	13.5%	0.84 [0.24, 2.90]
Total (95% CI)			**100.0%**	**0.63 [0.40, 1.00]**

Heterogeneity: Tau² = 0.00; Chi² = 0.23, df = 1 (P = 0.63); I² = 0%
Test for overall effect: Z = 1.96 (P = 0.05)



DYNAMIC BORROWING THROUGH SHRINKAGE ESTIMATION

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Röver & Friede (2020) SMMR

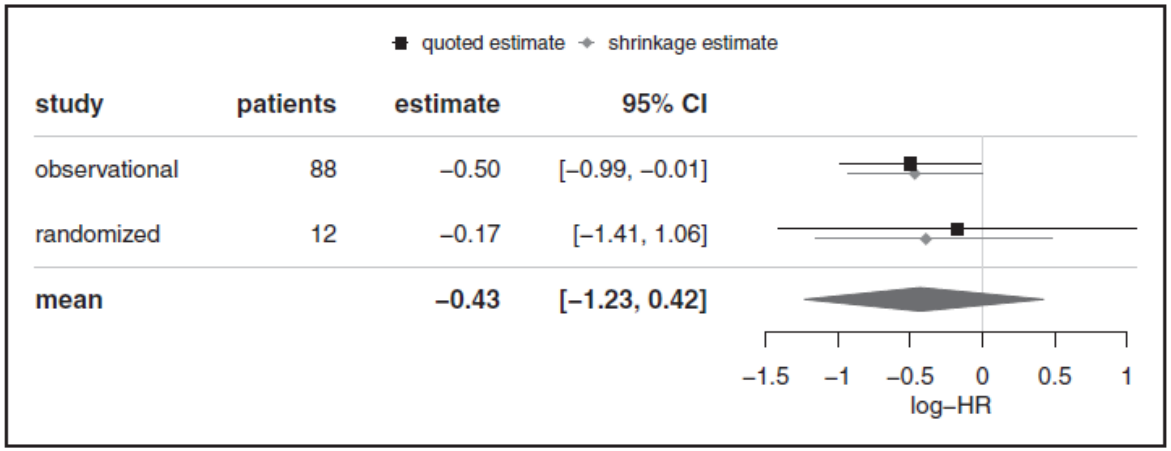


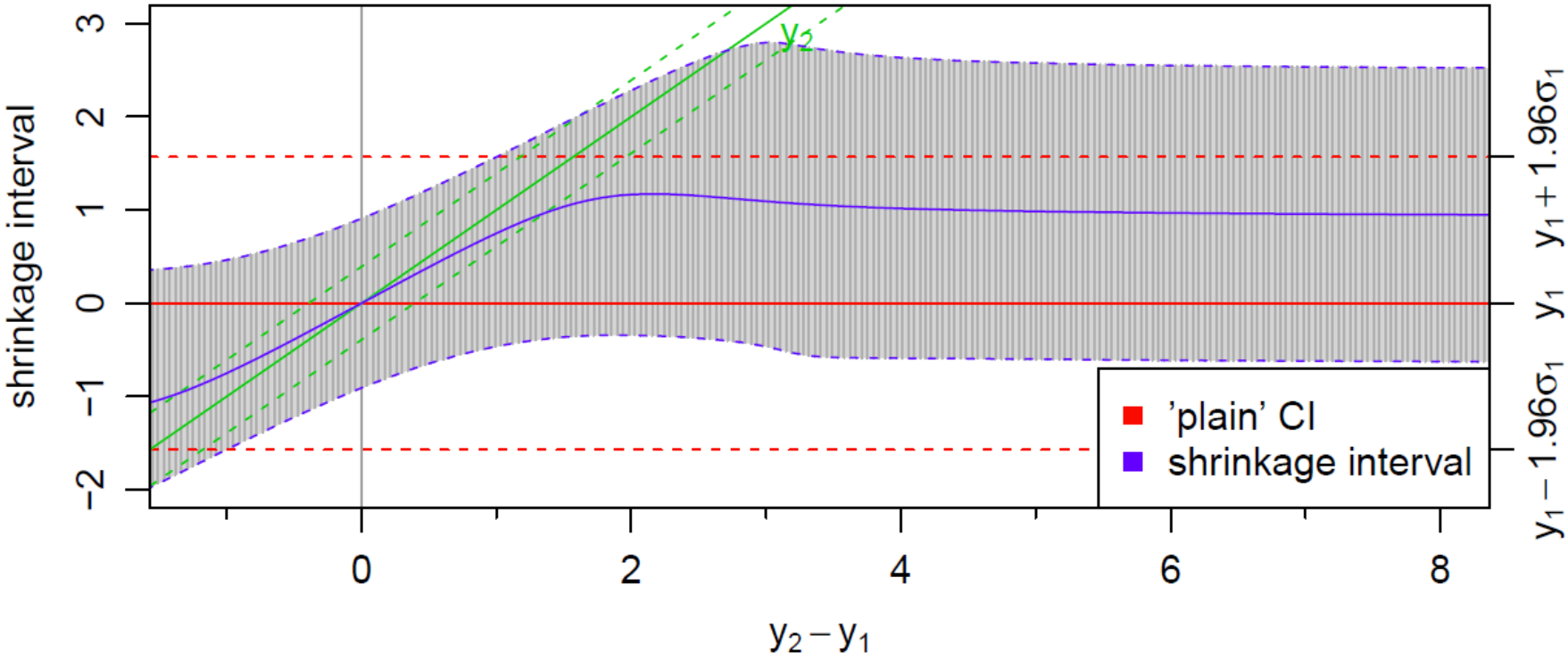
Figure 2. Forest plot for the CJD example (log-HR outcome). The shrinkage interval for the log-HR based on randomized evidence here is [−1.16, 0.48], spanning only two-thirds of the original confidence interval width.

- **Normal–normal hierarchical model** (NNHM)

$$y_i|\theta_i \sim \text{Normal}(\theta_i, s_i^2) \qquad \theta_i|\Theta, \tau \sim \text{Normal}(\Theta, \tau^2)$$

- **Bayesian framework**: Weakly informative prior on between-trial heterogeneity (R package **`bayesmeta`**)
- **RCT shrinkage interval**: 66% of original CI width; translates into 129% gain in sample size (about 27 instead of 12 patients)

SHRINKAGE ESTIMATION WITH K=2 STUDIES



- $n_1 = 25,\ n_2 = 400,\ p(\tau) = \text{HN}(0.5)$, interested in θ_1

Röver & Friede (2020) SMMR

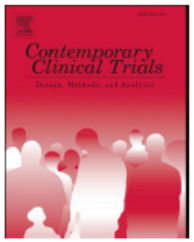
CAUSAL INFERENCE IN SMALL SAMPLES

Contemporary Clinical Trials 99 (2020) 106213

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial



Check for updates

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, Humboldtallee 32, 37073 Göttingen, Germany

ARTICLE INFO

Keywords:
COVID-19

ABSTRACT

The usual development cycles are too slow for the development of vaccines, diagnostics and treatments in pandemics such as the ongoing SARS-CoV-2 pandemic. Given the pressure in such a situation, there is a risk that

CONCLUSIONS AND DISCUSSION

- **Personalized medicine** (or stratified medicine)
- **Increasing number of applications of AI/ML** in clinical research and health care
- **Is there a role for statistics in AI / ML? Yes!**
  - Statistics contributes to **methods**, but impact on **applications** in my view more important
- **Synthesis of different data types** (e.g. RCT and clinical registries) as individual data sets often to small
- **Causal inference** increasingly important in clinical research (but better understanding of small sample properties needed)
- **Interdisciplinary networks** (such as at AICPM 2022 ☺)

EUROPEAN ASSOCIATION FOR DATA SCIENCE

ACKNOWLEDGEMENTS

- DZHK Project Group AI / ML

DZHK
DEUTSCHES ZENTRUM FÜR HERZ-KREISLAUF-FORSCHUNG E.V.

- DAGStat White Paper on Statistics in AI Working Party

DAG STAT
DEUTSCHE ARBEITSGEMEINSCHAFT STATISTIK

ANY QUESTIONS?

E-Mail: tim.friede@med.uni-goettingen.de
Homepage: https://medstat.umg.eu/
Twitter: @tim_friede

SOME REFERENCES

- Friedrich S, Antes G, Behr S, Binder H, Brannath W, Dumpert F, Ickstadt K, Kestler H, Lederer J, Leitgöb H, Pauly M, Steland A, Wilhelm A, Friede T (2021) Is there a role for statistics in artificial intelligence? Advances in Data Analysis and Classification. https://link.springer.com/article/10.1007/s11634-021-00455-6
- Friedrich S, Groß S, König IR, Engelhardt S, Bahls M, Heinz J, Huber C, Kaderali L, Kelm M, Leha A, Rühl J, Schaller J, Scherer C, Vollmer M, Seidler T, Friede T (2021) Applications of AI/ML approaches in cardiovascular medicine: A systematic review with recommendations. European Heart Journal - Digital Health 2: 424–436. https://academic.oup.com/ehjdh/advance-article/doi/10.1093/ehjdh/ztab054/6294933

SOME REFERENCES

- 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.
- Huber C, Benda N, Friede T (2022) Subgroup identification in individual participant data meta-analysis using model-based recursive partitioning. Advances in Data Analysis and Classification 16: 797–815.
- Röver C, Friede T (2020) Dynamically borrowing strength from another study through shrinkage estimation. Statistical Methods in Medical Research 29: 293–308.