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

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

Sarah Friedrich, Stefan Groß, Inke R. König, Sandy Engelhardt, Martin Bahls, Judith Heinz, Cynthia Huber, Lars Kaderali, Marcus Kelm, Andreas Leha, Jasmin Rühl, Jens Schaller, Clemens Scherer, Marcus Vollmer, Tim Seidler, and Tim Friede
**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
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

---

**Reproducibility**

- **Deficits**
  - 10% Trial registration
  - 90% No registration

- **Availability**
  - 8% direct access
    - 34% yes
    - 66% no
  - 14% code/model
    - 66% yes
    - 34% no
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/

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 Author(s) 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
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

- **Representativeness**
  - 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](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

- 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

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
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

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

Cynthia Huber¹ · Norbert Benda¹,² · Tim Friede¹

Received: 21 September 2020 / Revised: 25 June 2021 / Accepted: 22 July 2021 / Published online: 14 August 2021
© The Author(s) 2021
ADAPTIVE ENRICHMENT DESIGN

Interim analysis

Stage 1  Stage 2

F  S

Option

Futility stopping / Early success

F only

S only (Enrichment)

F and S

Friede et al. (2012) Stat Med
SIMULATING CLINICAL TRIALS

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

Tim Friede¹ | Nigel Stallard² | Nicholas Parsons²

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

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

Funding information
Medical Research Council, Grant/Award Number: G0001244

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.

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)

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
DYNAMIC BORROWING THROUGH SHRINKAGE ESTIMATION

Normal–normal hierarchical model (NNHM)

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

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

$n_1 = 25, \ n_2 = 400, \ p(\tau) = HN(0.5), \ \text{interested in } \theta_1$

Röver & Friede (2020) SMMR
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 😊)
The European Association for Data Science (EuADS) aims to provide a setting for fostering communication and cooperation among all stakeholders of data science. It is a platform for promoting the development and application of data science in various fields.
ACKNOWLEDGEMENTS

- DZHKO Project Group AI / ML
ANY QUESTIONS?

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


SOME REFERENCES

