Klinikum rechts der Isar
Technische Universität München

Statistische Analyse von Routinedaten in klinischer Forschung und Versorgungsforschung

24. SOMMERTAGUNG DES ARBEITSKREISES MEDIZINISCHER ETHIK-KOMMISSIONEN
Berlin, 13. Juni 2024

Alexander Hapfelmeier
Technical University of Munich
School of Medicine and Health
Institute of AI and Medical Informatics in Medicine

Grundlegende Prinzipien der klinischen Prüfung

ICH E9: Statistical principles for clinical trials[1]

"The most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be normal features of most controlled clinical trials intended to be included in a marketing application.

… it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects.

It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar."

[1]https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline

Quantitative Evaluation des Behandlungseffektes

Average treatment effect (ATE): $E_X(E(Y|T=1, X=x) - E(Y|T=0, X=x)))$

Bedingungen für kausale Inferenz[1]:

(1) Consistency: Alle behandelten Personen bekommen die gleiche Intervention.

(2) Positivity: Alle Personen kommen für die Intervention in Frage.

(3) Exchangeability: Der erwartete Effekt ist unabhängig von der Zuweisung der Intervention.

ATE in einem RCT direkt schätzbar: $\frac{1}{n_1}\sum y_{i1} - \frac{1}{n_0}\sum y_{i0}$

[1]Sourial N, Longo C, Vedel I, Schuster T. Daring to draw causal claims from non-randomized studies of primary care interventions. Fam Pract. 2018 Sep 18;35(5):639-643. doi: 10.1093/fampra/cmy005

Real-world data & Big data

FDA[1]: „Real-world data are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. Examples of RWD include data derived from electronic health records, medical claims data, data from product or disease registries, and data gathered from other sources (such as digital health technologies) that can inform on health status.“

EMA[2]: „'Big data' is a widely-used term without a commonly-accepted definition. The HMA-EMA Big Data Task Force defined big data as ‘extremely large datasets which may be complex, multi-dimensional, unstructured and heterogeneous, which are accumulating rapidly and which may be analysed computationally to reveal patterns, trends, and associations.

… Sources include real-world data (such as electronic health records, insurance claims data and data from patient registries), genomics, clinical trials, spontaneous adverse drug reaction reports, social media and wearable devices.“

[1]https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence - accessed 27.03.2024
[2]https://www.ema.europa.eu/en/about-us/how-we-work/big-data

Potentiale, Erwartungen & Hoffnungen

Herausforderungen

Hohe Fallzahl (n) und große Anzahl erhobener Merkmale (p) ermöglichen/führen zu:

- Präzise Schätzung multipler Effekte
- Untersuchung komplexer Zusammenhänge
- mittels komplexer Analysemethoden (z.B. AI/ML)
- Subgruppenanalysen
- Methoden der personalisierten Medizin

- Falsch-positive Effekte
- Irrelevante Effekte
- Präzisionsverlust durch zusätzliches Rauschen

→ Nadel-im-Heuhaufen-Problem

- Verzerrte Effekte

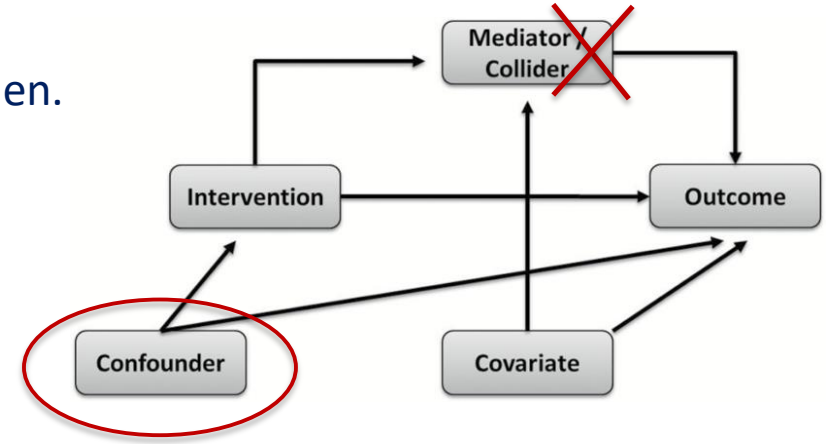
Real-world evidence

(3) Exchangeability: Der erwartete Effekt ist unabhängig von der Zuweisung der Intervention.

Problem: Unmeasured Confounding

Theorie: Durch geeignetes Bedingen auf Charakteristika X kann der erwartete Effekt unabhängig von der Zuweisung geschätzt werden.

Strongly ignorable treatment assignment (SITA)[1]: $T \perp \{Y(1), Y(0)\} \mid X$



Herstellen der Vergleichbarkeit durch[2]:

- Matching, Propensity Score (PS) Matching
- Inverse Probability of Treatment Weighting (IPTW)
- Stratifizierte Analyse
- Multiple Regressionsmodelle
- Methoden der kausalen Inferenz – „causal methods are the future“? [3]

[1]Paul R Rosenbaum, Donald B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika, Volume 70, Issue 1, April 1983, Pages 41–55, https://doi.org/10.1093/biomet/70.1.41
[2]Kuss O, Blettner M, Börgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl Int. 2016 Sep 5;113(35-36):597-603. doi: 10.3238/arztebl.2016.0597.
[3]Sourial N, Longo C, Vedel I, Schuster T. Daring to draw causal claims from non-randomized studies of primary care interventions. Fam Pract. 2018 Sep 18;35(5):639-643. doi: 10.1093/fampra/cmy005

„The Magic of Randomization versus the Myth of Real-World Evidence[1]"

Statins and the risk of colon cancer[2]

Background: … Statins inhibit the growth of colon-cancer cell lines, and secondary analyses of some, but not all, clinical trials suggest that they reduce the risk of colorectal cancer.

Methods: … a population-based case-control study of patients who received a diagnosis of colorectal cancer in northern Israel between 1998 and 2004 and controls matched according to age, sex, clinic, and ethnic group.

Table 2. Crude and Adjusted Associations between Statin Use and the Risk of Colorectal Cancer in the Unmatched Study Population.*

Variable	Patients	Controls	Total No.	Unadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)†
	no. of subjects (%)				
Statin use				0.50 (0.40–0.63)	0.57 (0.44–0.73)
Yes	120 (6.1)	234 (11.6)	354		
No	1833 (93.9)	1781 (88.4)	3614		

* CI denotes confidence interval.
† Analyses were adjusted for age; sex; use or nonuse of aspirin or other NSAIDs; ethnic group; the presence or absence of sports participation, hypercholesterolemia, and a history of colorectal cancer in a first-degree relative; and level of vegetable consumption.

Interpretation of the evidence for the efficacy and safety of statin therapy (RCTs)[3]

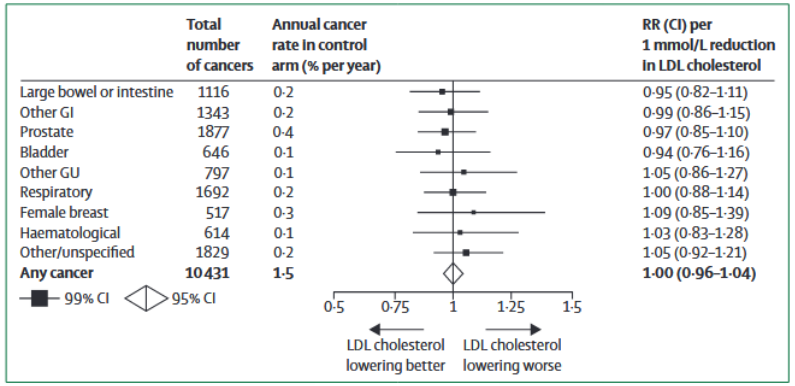


***Figure 7*: Effects of lowering LDL cholesterol with statin therapy on site-specific cancer in meta-analyses of randomised trials of statin therapy**
Adapted from CTT Collaboration website. Combined comparisons in randomised trials of routine statin therapy versus no routine statin therapy and of more versus less intensive statin therapy. RR=rate ratio. GI=gastrointestinal. GU=genitourinary.

Kritik and real-world evidence[1]:

- Verzerrte Effektschätzung (Bias)
- Keine sichere Adjustierung möglich

[1]Collins, R., Bowman, L., Landray, M., & Peto, R. (2020). The magic of randomization versus the myth of real-world evidence. New England Journal of Medicine, 382(7), 674-678.
[2]Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G. Statins and the risk of colorectal cancer. N Engl J Med. 2005 May 26;352(21):2184-92. doi: 10.1056/NEJMoa043792
[3]Collins, R., Reith, C., Emberson, J., Armitage, J., Baigent, C., Blackwell, L., … & Peto, R. (2016). Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet, 388(10059), 2532-2561.

Überleben nach Herzinfarkt

Sex differences in long-term mortality among acute myocardial infarction patients: Results from the ISAR-RISK and ART studies

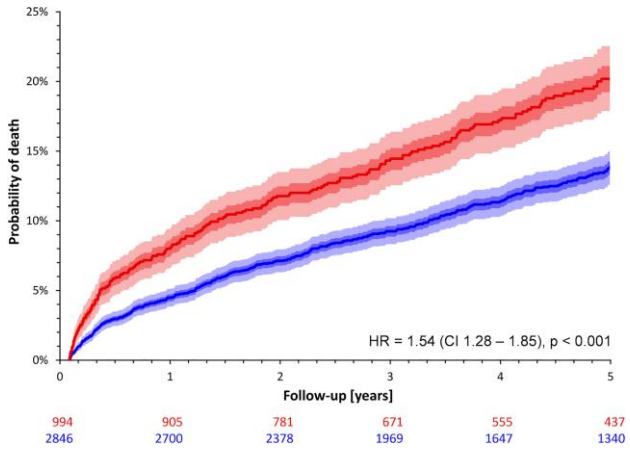
Background: Mortality rates in females who survived acute myocardial infarction (AMI) exceed those in males. Differences between sexes in age, cardiovascular risk factors and revascularization therapy have been proposed as possible reasons.

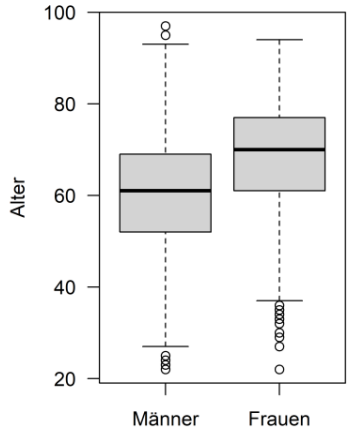
Objective: To select sets of female and male patients comparable in respect of relevant risk factors in order to compare the sex-specific risk in a systematic manner.

Methods: Data of the ISAR-RISK and ART studies were investigated [Cohort Studies]. Patients were enrolled between 1996 and 2005 and suffered from AMI within 4 weeks prior to enrolment. Patients of each sex were selected with 1:1 equivalent age, previous AMI history, sinus-rhythm presence, hypertension, diabetes mellitus, smoking status, left ventricular ejection fraction (LVEF), and revascularization therapy. Survival times were compared between sex groups in the whole study cohort and in the matched cohort.

Ubrich, R., Barthel, P., Haller, B., Hnatkova, K., Huster, K. M., Steger, A., ... & Schmidt, G. (2017). Sex differences in long-term mortality among acute myocardial infarction patients: Results from the ISAR-RISK and ART studies. PloS one, 12(10), e0186783.

Überleben nach Herzinfarkt – Ergebnisse





Matching

Fig 1. Probabilities of death stratified by sex in the entire study cohort.

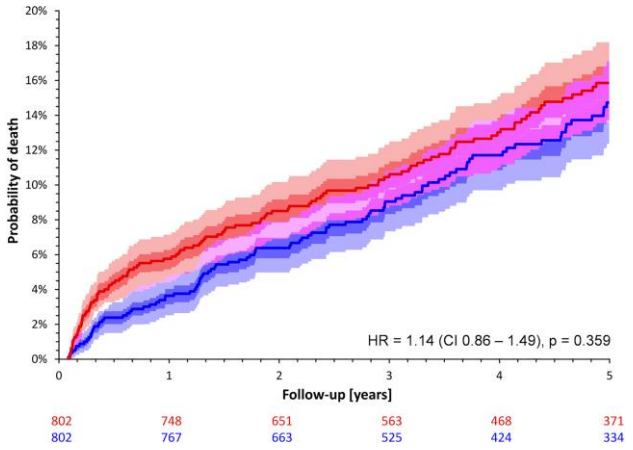


Fig 2. Probabilities of death stratified by sex in the cohort of matched patients.

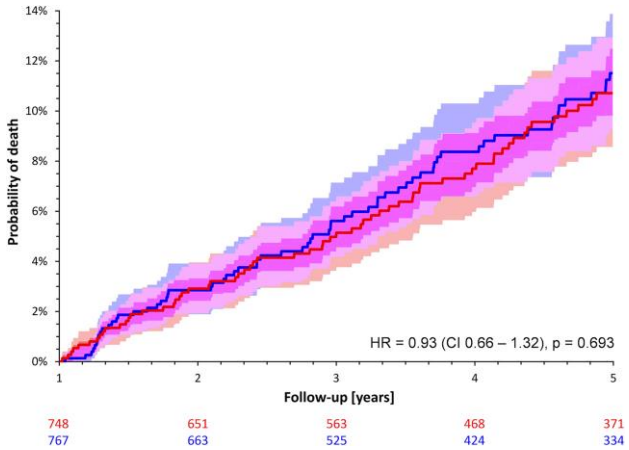


Fig 4. Probabilities of death stratified by sex in the sex-matched cohort between the second and fifth year after the index infarction.

Ubrich, R., Barthel, P., Haller, B., Hnatkova, K., Huster, K. M., Steger, A., ... & Schmidt, G. (2017). Sex differences in long-term mortality among acute myocardial infarction patients: Results from the ISAR-RISK and ART studies. PloS one, 12(10), e0186783.

„Where Medical Statistics meets Artificial Intelligence“ (NEJM)[1]:

Causal ML for high dimensional confounding

- Hard to interpret
- Instable
- Reproducibility
- Garbage in garbage out

Deshalb:

- Sorgsames prüfen der Datenqualität
- Transparente Beschreibung der Methoden
- Ergebnisse bereitstellen, kritisch hinterfragen, Sensitivitätsanalysen, Kontext
- Notwendigkeit interner und externer Validierung

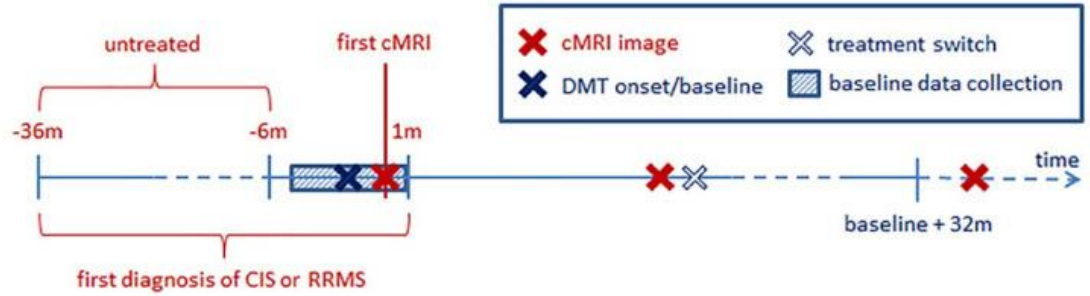
[1]Hunter, D. J., & Holmes, C. (2023). Where medical statistics meets artificial intelligence. New England Journal of Medicine, 389(13), 1211-1219.

Beispiel klinische Forschung / prädiktive Modellierung

Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis[1]

Datenquellen: Klinische Daten, Bildgebung (MRI), Labordaten (n: 2992 → 475, p: 65)

Datenextraktion:



Analyse: A predictive RF … on fully parameterized Cox proportional hazards models … had treatment (no medication versus platform medication) as a predictor variable in the underlying Cox models while other features were used … to build the tree structure … the interaction of the features with treatment is explicitly modelled.

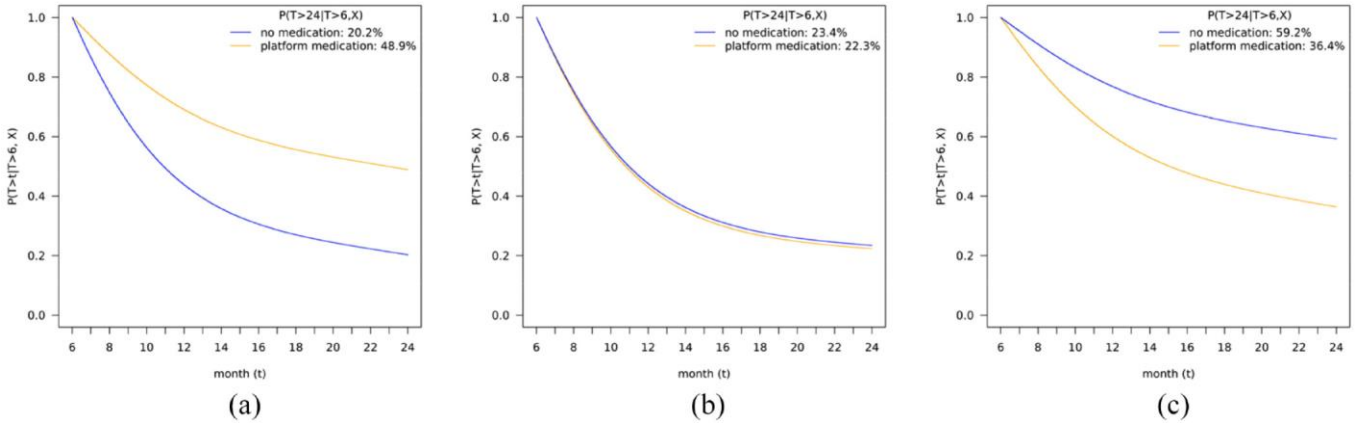
[1]Hapfelmeier, Alexander, et al. "Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis." Therapeutic Advances in Neurological Disorders 16 (2023): 17562864231161892.

Beispiel klinische Forschung – Ergebnisse

Ergebnis:

- Intern validierte AUC = 0.624 („considered weak but is comparable to … other models …“)
- 40 relevante Prädiktoren

Patientenindividuelle Prädiktion:



Limitationen[1]:

- „underlying assumption … might not hold. This might result in biased effect estimation and models.“
- „The robustness … has yet to be demonstrated in a prospective study … for external validation.“[2]

[1]Hapfelmeier, Alexander, et al. "Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis." Therapeutic Advances in Neurological Disorders 16 (2023): 17562864231161892.
[2]Bayas, Antonios, et al. "Prospective study validating a multidimensional treatment decision score predicting the 24-month outcome in untreated patients with clinically isolated syndrome and early relapsing–remitting multiple sclerosis, the ProVal-MS study." Neurological Research and Practice 6.1 (2024): 15.

Beispiel Versorgungsforschung / Effektschätzung

Forschungsfrage: Zusammenhang leitliniengerechte Medikationsadhärenz (Statine, BB, ASS etc.) und Kosten (amb., stat., med.) am Beispiel KHK (n=100.157 GKV-Versicherte)[1,2]

Freiheitsgrade im Umgang mit Datenaufbereitung und Analysemethoden:

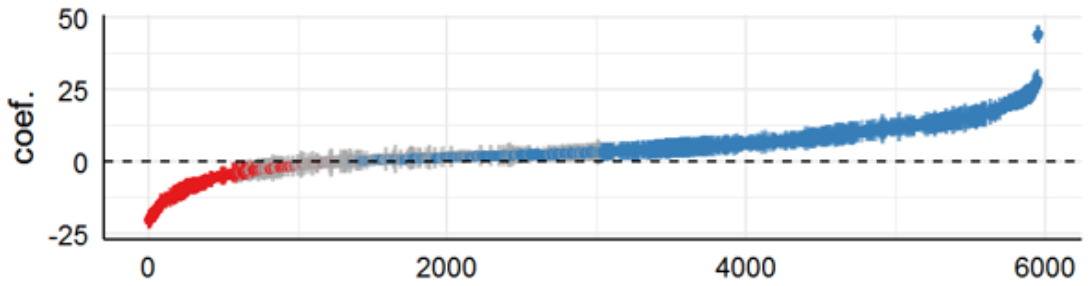
- Definition Studienpopulation (chronisch Kranke, Medikation, Ausreißer „schwer Kranke“)
- Definition Adhärenz (KH-Aufenthalte, Carry-Over, Polypharmazie)
- Definition Model (zeitverzögerte Effekte, Umgang mit Störgrößen)

→ 5952 Möglichkeiten

Wie verlässlich/robust sind die Ergebnisse einer Analyse von GKV-Routinedaten?

→ "vibration of effects" | "multiverse analysis" |

Kritik: Unterschiedliche estimands!
(ICH E9 addendum)



[1]Wendl J, Simon A, Kistler M, Hapfelmeier J, Schneider A and Hapfelmeier A (2022), Identification of target groups and individuals for adherence interventions using treebased prediction models. Front. Pharmacol. 13:1001038. doi: 10.3389/fphar.2022.1001038
[2]Wendl, J., Simon, A., Kistler, M., Hapfelmeier, J., Schneider, A., & Hapfelmeier, A. (2023). Medication Adherence and Healthcare Costs in Chronically Ill Patients Using German Claims Data. Applied Health Economics and Health Policy, 21(3), 477-487.

Zusammenfassung

Analysen von Real-world data

- haben großes Potential für die Generierung von Evidenz.
- Es existieren theoretisch fundierte Analysemethoden.
- In der Praxis kann eine Verzerrung von Ergebnissen nicht ausgeschlossen werden.
- Evidenz aus experimentellen Studien (RCT) ist weiterhin erforderlich.
- Umsichtige und transparente Datenverarbeitung, Analyse und Interpretation ist geboten!

Vielen Dank für die Aufmerksamkeit!

Literatur

- Sourial N, Longo C, Vedel I, Schuster T. Daring to draw causal claims from non-randomized studies of primary care interventions. Fam Pract. 2018 Sep 18;35(5):639-643.
- Paul R Rosenbaum, Donald B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika, Volume 70, Issue 1, April 1983
- Kuss O, Blettner M, Börgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl Int. 2016 Sep 5;113(35-36):597-603.
- Ubrich, R., Barthel, P., Haller, B., Hnatkova, K., Huster, K. M., Steger, A., ... & Schmidt, G. (2017). Sex differences in long-term mortality among acute myocardial infarction patients: Results from the ISAR-RISK and ART studies. PloS one, 12(10), e0186783.
- Collins, R., Bowman, L., Landray, M., & Peto, R. (2020). The magic of randomization versus the myth of real-world evidence. New England Journal of Medicine, 382(7), 674-678.
- Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G. Statins and the risk of colorectal cancer. N Engl J Med. 2005 May 26;352(21):2184-92.
- Collins, R., Reith, C., Emberson, J., Armitage, J., Baigent, C., Blackwell, L., ... & Peto, R. (2016). Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet, 388(10059), 2532-2561.
- Hunter, D. J., & Holmes, C. (2023). Where medical statistics meets artificial intelligence. New England Journal of Medicine, 389(13), 1211-1219.
- Hapfelmeier, Alexander, et al. "Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis." Therapeutic Advances in Neurological Disorders 16 (2023): 17562864231161892.
- Bayas, Antonios, et al. "Prospective study validating a multidimensional treatment decision score predicting the 24-month outcome in untreated patients with clinically isolated syndrome and early relapsing–remitting multiple sclerosis, the ProVal-MS study." Neurological Research and Practice 6.1 (2024): 15.
- Wendl J, Simon A, Kistler M, Hapfelmeier J, Schneider A and Hapfelmeier A (2022), Identification of target groups and individuals for adherence interventions using treebased prediction models. Front. Pharmacol. 13:1001038.
- Wendl, J., Simon, A., Kistler, M., Hapfelmeier, J., Schneider, A., & Hapfelmeier, A. (2023). Medication Adherence and Healthcare Costs in Chronically Ill Patients Using German Claims Data. Applied Health Economics and Health Policy, 21(3), 477-487.