Randomisierte vs. nichtrandomisierte Studien: Evidenz aus der Meta-Epidemiologie

Oliver Kuß

Institut für Biometrie und Epidemiologie, Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität Düsseldorf

und

Centre for Health and Society (chs), Medizinische Fakultät der Heinrich-Heine-Universität Düsseldorf

Limitationen von RCTs

DDZ German Diabetes Center

- Limitierte Verallgemeinerbarkeit (externe Validität)
  RCTs werden häufig in selektionierten (i.a. gesünderen) Populationen durchgeführt, die nicht repräsentativ für den durchschnittlichen Patienten sind.
- Limitierte Relevanz für die Versorgung ("Efficacy-Effectiveness-Gap")
  RCTs überschätzen in der Regel den Effekt der Behandlung in der Routine-Anwendung.
- Limitierte Aussagen zu Sicherheit und Nebenwirkungen

Frieden TR. Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. N Engl J Med. 2017 Aug 3;377(5):465-475.
Rawlins M (2008). *De Testimonio: on the evidence for decisions about the use of therapeutic interventions*. Royal College of Physicians. Available at: http://www.amcp.org/WorkArea/DownloadAsset.aspx?id=12451

Limitationen von RCTs

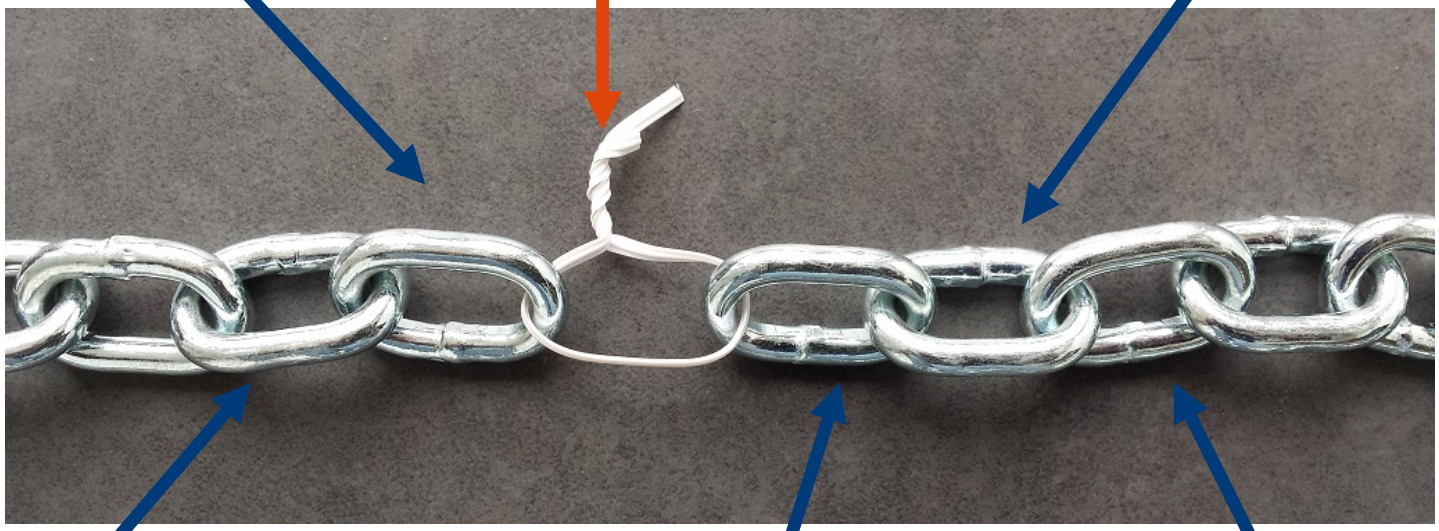
DDZ German Diabetes Center

- Hoher Ressourcenbedarf
  RCTs sind in der Regel teuer, zum Teil verursacht durch regulatorische Auflagen, die die Teilnehmenden schützen sollen. Es besteht zusätzliche Gefahr, dass deshalb Surrogat-Marker verwendet werden oder dass Studien zu klein sind, um Effektheterogenität abzubilden (Subgruppen-Analysen).
- Lange Dauer
  Planung, Durchführung und Analyse von RCTs dauern in der Regel Jahre, so dass RCTs Gefahr laufen, mit dem Tempo der klinischen Innovation nicht Schritt zu halten.
- Ethische Aspekte
  Es gibt Situationen, wo RCTs „unnötig, ungeeignet, unmöglich oder ungenügend“ (Black, 1996) sind.

Black N: Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996; 312:1215–8.

Nichtrandomisierte Studien

DDZ German Diabetes Center



Externe Validität

Fehlende Randomisierung (Interne Validität)

Geringer Ressourcenbedarf

Hohe Relevanz für die Versorgung

Aussagen zur Sicherheit möglich

Geschwindigkeit

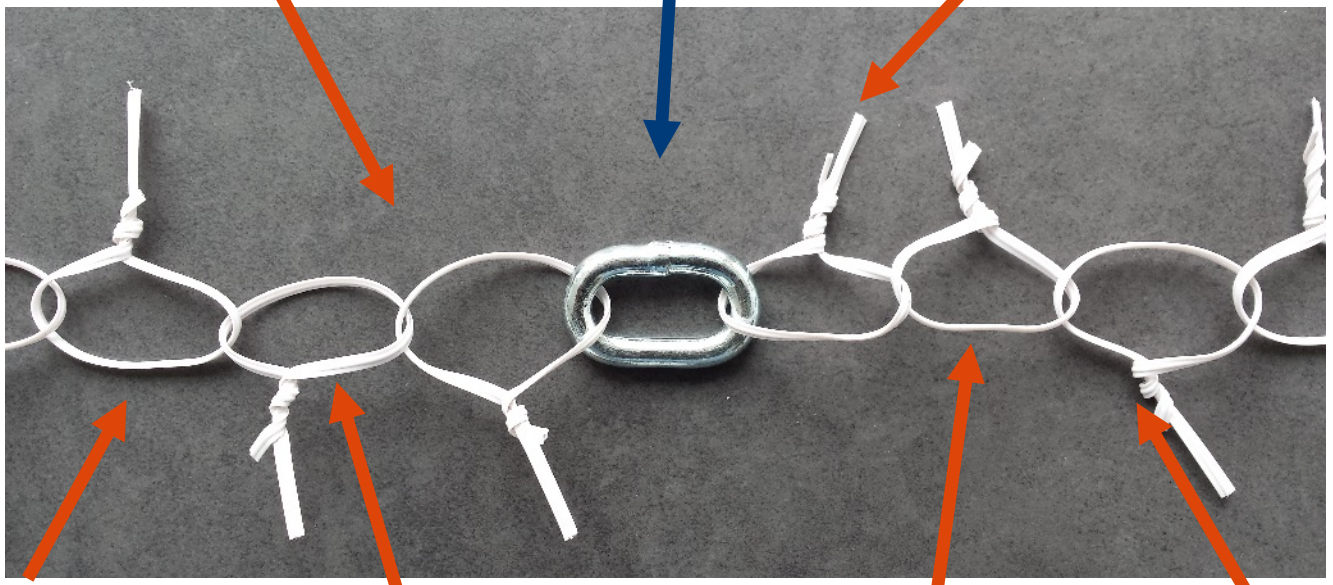
Randomisierte Studien

DDZ German Diabetes Center

Limitierte Verallgemeinerbarkeit

Randomisierung (Interne Validität)

Limitierte Relevanz für die Versorgung



Limitierte Aussagen zur Sicherheit

Hoher Ressourcenbedarf

Lange Dauer

Ethische Herausforderungen

Designs zum Vergleich von RCTs und Non-RCTs

DDZ
German Diabetes Center

- **Grundsätzliches Problem**:

  Wenn **RCTs** in **selektionierten** Populationen durchgeführt werden, **Non-RCTs** dagegen in **unselektionierten**, dann sind Unterschiede nicht notwendigerweise auf die **fehlende Randomisierung** zurückzuführen.
  Sie könnten auch durch die **unterschiedlichen Patientenpopulationen** zustande kommen!

Designs zum Vergleich von RCTs und Non-RCTs

German Diabetes Center

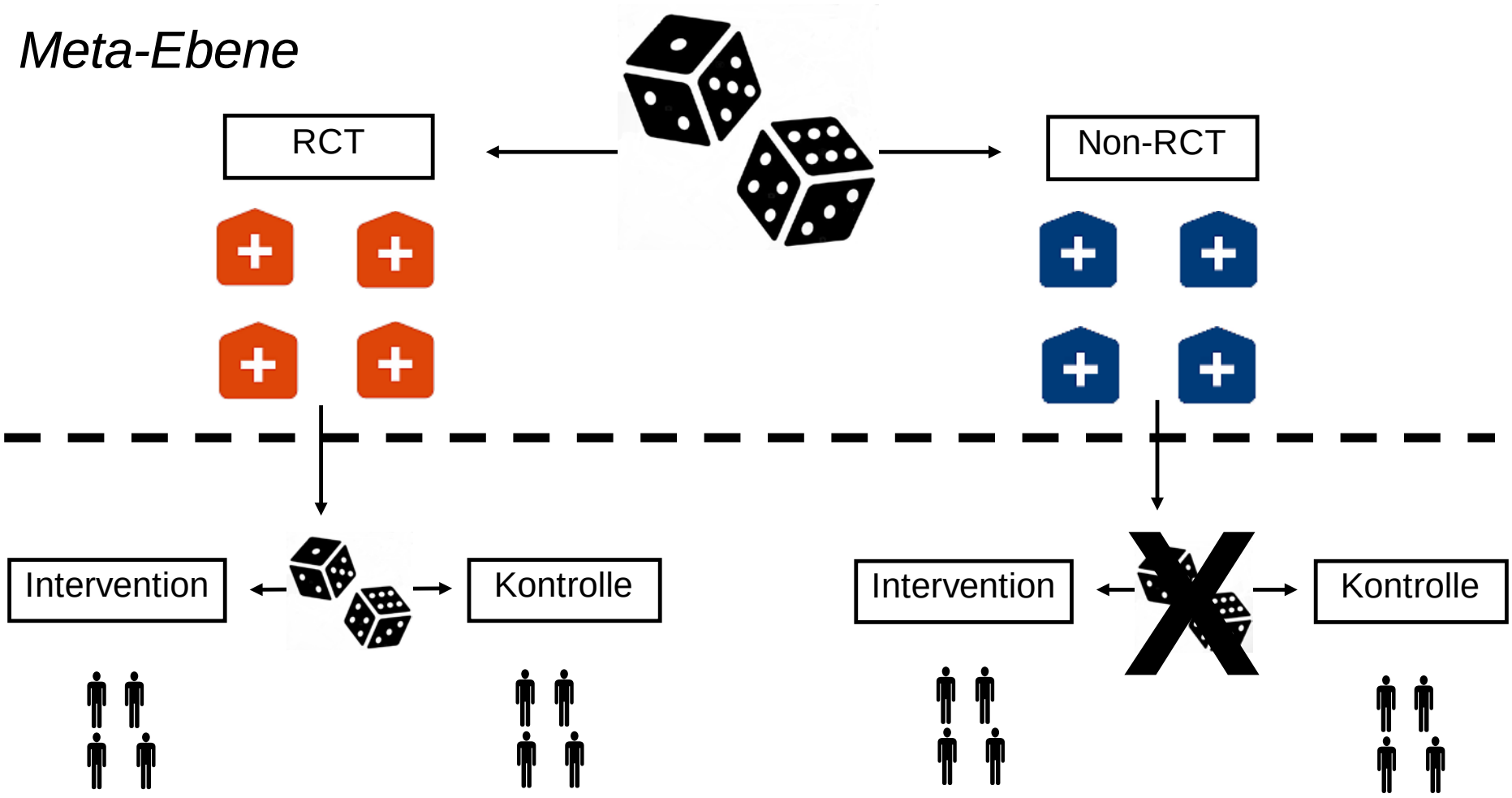
- **Lösung: „Meta-Randomisierung“**
  Studiengruppen, die eine Studie zu einer bestimmten klinischen Frage durchzuführen bereit sind, werden zufällig ausgewählt („meta-randomisiert“), eine randomisierte oder eine nichtrandomisierte Studie durchzuführen.

  → Strukturgleichheit bzgl. aller „Meta-Confounder“
  → Kausaler Effekt der Randomisierung ist intern valide messbar.

- Technisch möglich? Ethisch akzeptabel?

Meta-Randomisierung

DDZ German Diabetes Center



Designs zum Vergleich von RCTs und Non-RCTs

German Diabetes Center

- Auf individueller Ebene wurden meta-randomisierte Studien bei der Untersuchung von Patienten-Präferenzen durchgeführt („Doubly Randomized Preference Trials“).
- Systematischer Review von Studien (mit und ohne Meta-Randomisierung):

“Differences in outcome across the trials between randomized and preference groups were generally small, particularly in large trials and after accounting for baseline measures of outcome. Therefore, there was little evidence that preferences substantially interfere with the internal validity of randomized trials. […] Preferences influence whether people participate in randomized trials, but there is little evidence that they significantly affect validity.

King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA. 2005 Mar 2;293(9):1089-99.

A doubly randomized preference trial

German Diabetes Center

Clinical Care/Education/Nutrition

ORIGINAL ARTICLE

Patient Choice in Diabetes Education Curriculum

Nutritional versus standard content for type 2 diabetes

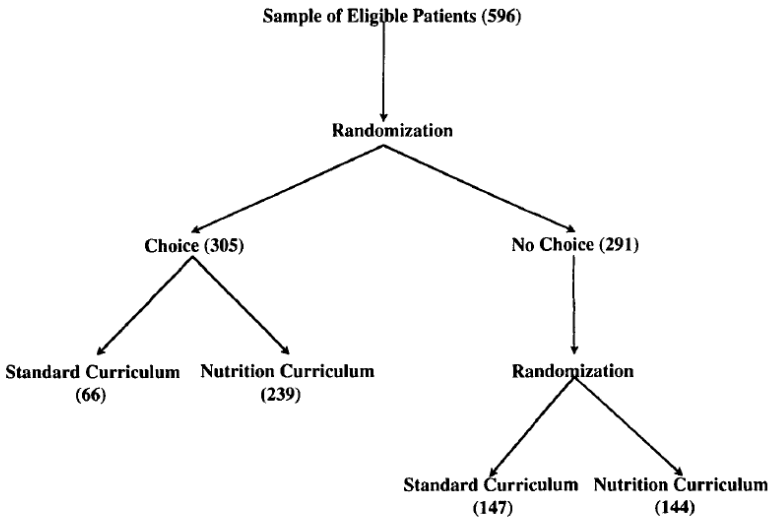


Figure 1—*Study design.*

Noël PH, Larme AC, Meyer J, Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. Nutritional versus standard content for type 2 diabetes. Diabetes Care. 1998 Jun;21(6):896-901.

A doubly randomized preference trial

DDZ
German Diabetes Center

"Contrary to our hypothesis, patients who were allowed to choose their curriculum did not have significantly higher attendance rates or significantly better improvements in diabetes knowledge or other clinical outcomes compared with patients who were randomly assigned to the two different curriculum types."

Noël PH, Larme AC, Meyer J, Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. Nutritional versus standard content for type 2 diabetes. Diabetes Care. 1998 Jun;21(6):896-901.

Within-study comparisons

DDZ German Diabetes Center

- In den Sozialwissenschaften heißen Studien zum Vergleich von randomisierten und nicht-randomisierten Studien "Within-study comparisons".
- Beispiel mit Meta-Randomisierung: Vergleich Mathematik- oder Vokabel-Nachhilfe vor Test bei Psychologie-Studierenden in den USA (Shadish et al., 2008)…

  "This study suggests that adjusted results from nonrandomized experiments can approximate results from randomized experiments. […] All of the recommended adjustments always reduced bias (and never increased it), and did so substantially."
- … mit unabhängiger Replikation in Deutschland (Pohl et al., 2009)

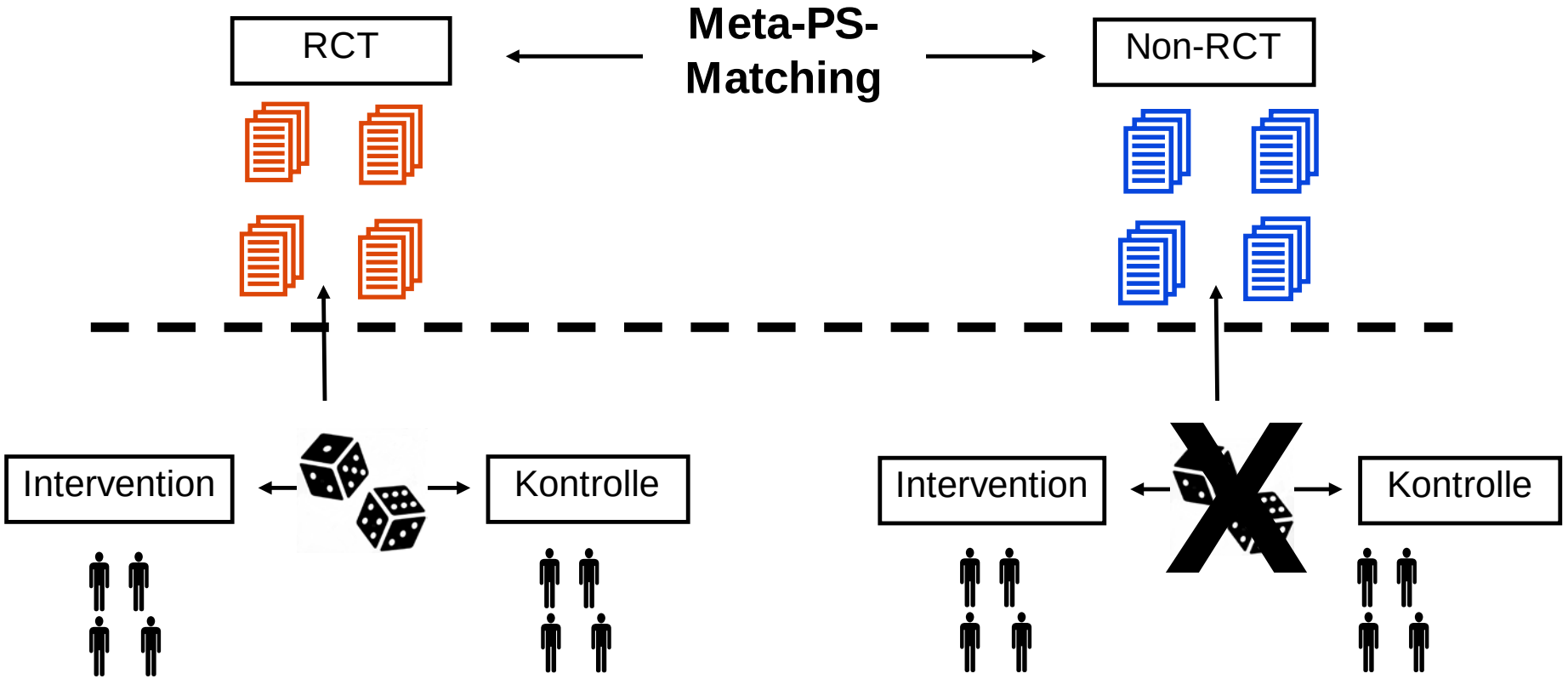
Shadish WR, Clark MH, Steiner PM. Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. Journal of the American Statistical Association. 2008 Dec;103(484):1334-1343.

Pohl S, Steiner PM, Eisermann J, Soellner R, Cook TD. Unbiased Causal Inference From an Observational Study: Results of a Within-Study Comparison. Educational Evaluation and Policy Analysis. 2009 Dec;31(4):463–479.

Meta-Propensity Score-Matching

German Diabetes Center

Meta-Ebene

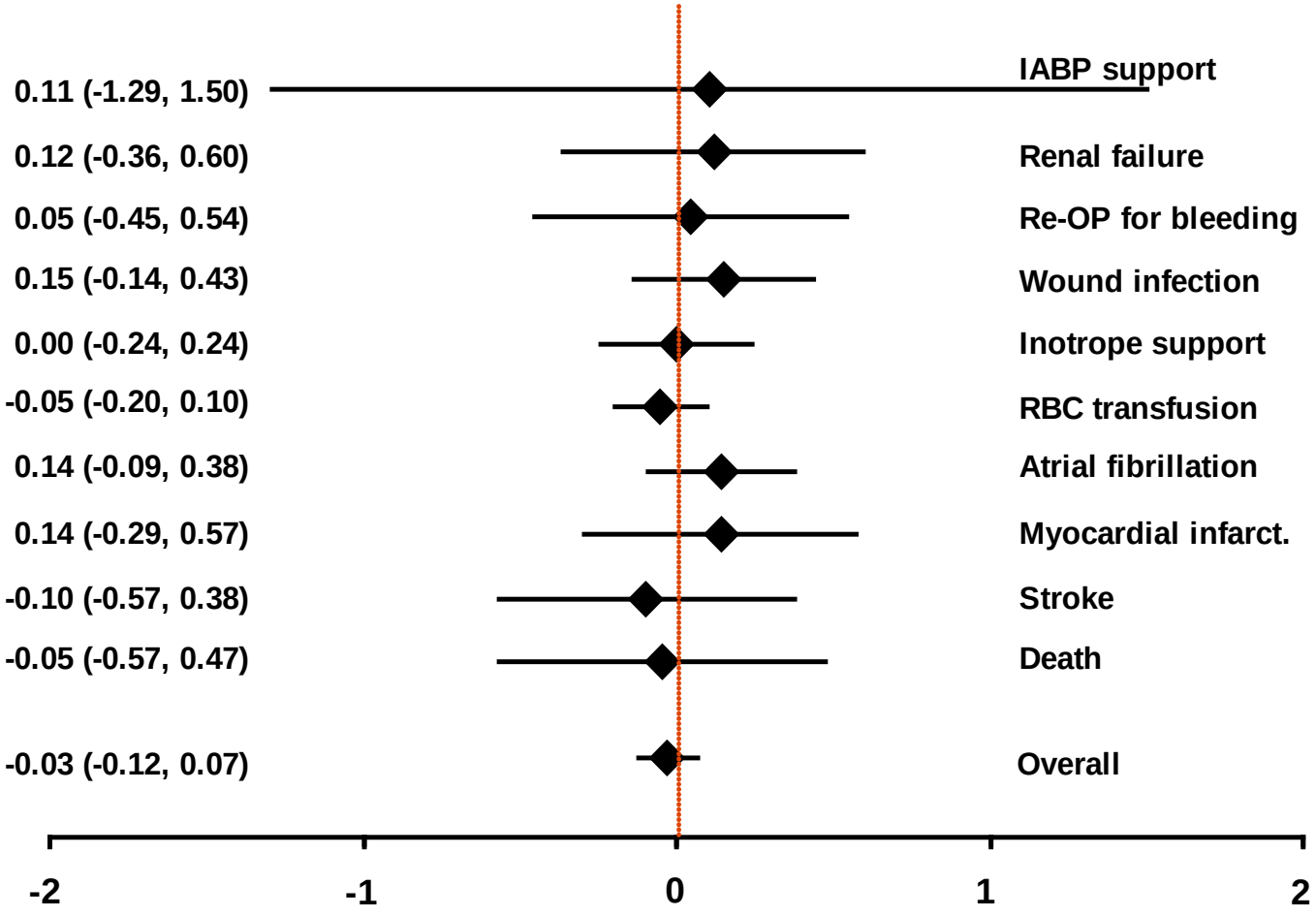


Inhaltlich-Klinische Ebene

Kuss O, Legler T, Börgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. J Clin Epidemiol. 2011 Oct;64(10):1076-84.

Differenzen der „Meta-ORs“(PS-RCT) im „meta-gematchten“ Datensatz

German Diabetes Center



Kuss O, Legler T, Börgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. J Clin Epidemiol. 2011 Oct;64(10):1076-84.

Die Mutter aller systematischen Reviews

German Diabetes Center

Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials (Review)

Anglemyer A, Horvath HT, Bero L

THE COCHRANE COLLABORATION®

Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):MR000034.

Die Mutter aller systematischen Reviews

DDZ German Diabetes Center

OBJECTIVES:
To assess the impact of study design (including RCTs versus observational study designs) on the effect measures estimated.

MAIN RESULTS:
Our initial search yielded 4406 unique references. Fifteen reviews met our inclusion criteria; 14 of which were included in the quantitative analysis. The included reviews analyzed data from 1583 meta-analyses that covered 228 different medical conditions. […] Our primary quantitative analysis, including 14 reviews, showed that the pooled ROR comparing effects from RCTs with effects from observational studies was 1.08 (95% confidence interval (CI) 0.96 to 1.22). Of 14 reviews included in this analysis, 11 (79%) found no significant difference between observational studies and RCTs. One review suggested observational studies had larger effects of interest, and two reviews suggested observational studies had smaller effects of interest.

AUTHORS' CONCLUSIONS:
Our results across all reviews (pooled ROR 1.08) are very similar to results reported by similarly conducted reviews. As such, we have reached similar conclusions; **on average, there is little evidence for significant effect estimate differences between observational studies and RCTs, regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions.**

Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):MR000034.

Target trial emulation

DDZ German Diabetes Center

- **Definition** (Labrecque/Swanson, 2017): Anwendung der Design-Prinzipien aus RCTs für die Analyse von Beobachtungsdaten mit expliziter Verknüpfung zum RCT, der emuliert werden soll.
- Komponenten, die spezifiziert werden müssen (Hernán/Robins, 2016): Ein-/Ausschlusskriterien, Behandlungsstrategien, Therapiezuweisung, Länge Follow-Up, Zielgrößen, Kausale Kontraste, Analyseplan

Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. Eur J Epidemiol. 2017 Jun;32(6):473-475.
Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64.

Target trial emulation

German Diabetes Center

- **Nutzen hier:** Vergleiche Non-RCTs mit den RCTs, die emuliert werden.
- **Prominentes Beispiel:** Emulation der Women's Health Initiative (WHI)-Studie zur Hormonersatztherapie bei postmenopausalen Frauen.

"Our findings suggest that the discrepancies between the Women's Health Initiative and Nurses' Health Study ITT estimates could be largely explained by differences in the distribution of time since menopause and length of follow-up."

Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008 Nov;19(6):766-79.

Target trial emulation

DDZ
German Diabetes Center

- **Klar:** RCT-Resultate sollten für die Emulierenden verblindet sein.
- **Beispiel** (Admon et al. 2019):

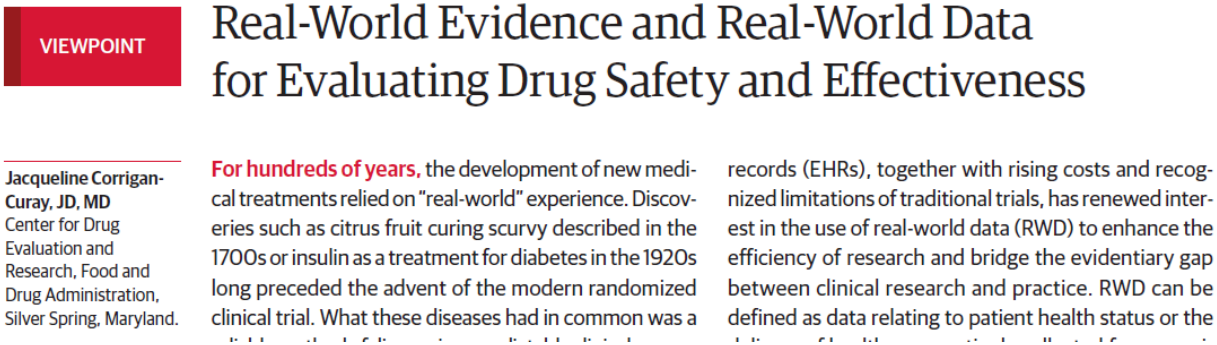
> OBJECTIVES: We tested the hypothesis that blinded analysts applying target trial emulation to existing observational data could predict the results of an RCT.
>
> CONCLUSIONS: Applying target trial emulation methods to existing observational data for the evaluation of a novel intervention produced results similar to those of a randomized trial. These findings support the use of target trial emulation for comparative effectiveness research.

Admon AJ, Donnelly JP, Casey JD, Janz DR, Russell DW, Joffe AM, Vonderhaar DJ, Dischert KM, Stempek SB, Dargin JM, Rice TW, Iwashyna TJ, Semler MW; Pragmatic Critical Care Research Group. Emulating a Novel Clinical Trial Using Existing Observational Data: Predicting Results of the PreVent Study. Ann Am Thorac Soc. 2019 Apr 30

Target trial emulation

DDZ German Diabetes Center

[…] Under the 21st Century Cures Act, the Food and Drug Administration is tasked with developing a program to evaluate the use of RWE to support approval of new indications for approved drugs or to satisfy postapproval study requirements. […]



VIEWPOINT

Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness

Jacqueline Corrigan-Curay, JD, MD
Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland.

For hundreds of years, the development of new medical treatments relied on "real-world" experience. Discoveries such as citrus fruit curing scurvy described in the 1700s or insulin as a treatment for diabetes in the 1920s long preceded the advent of the modern randomized clinical trial. What these diseases had in common was a

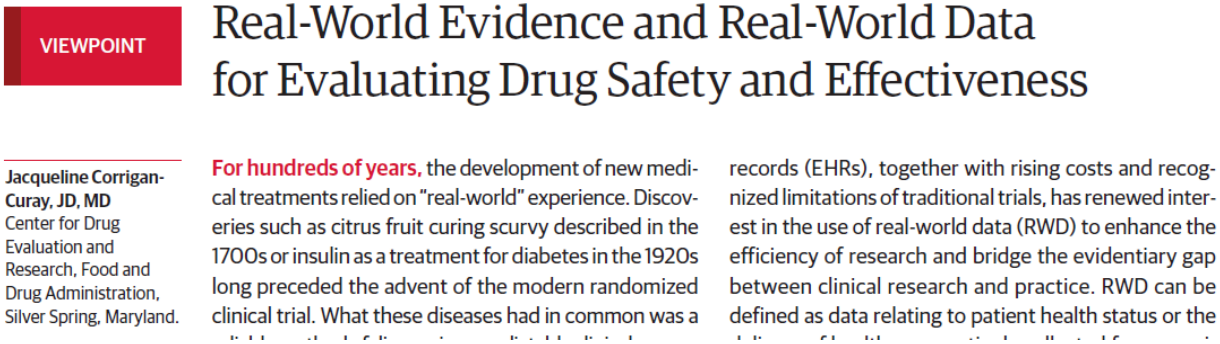
records (EHRs), together with rising costs and recognized limitations of traditional trials, has renewed interest in the use of real-world data (RWD) to enhance the efficiency of research and bridge the evidentiary gap between clinical research and practice. RWD can be defined as data relating to patient health status or the

Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA. 2018 Sep 4;320(9):867-868.

Target trial emulation

DDZ German Diabetes Center

„[…] the FDA is funding a study to explore whether observational methods can be used to replicate the results of approximately 30 clinical trials designed to provide evidence about the effectiveness of a drug. This project will assist the FDA in understanding how observational methods can be applied to address questions involving drug effectiveness.” (Corrigan-Curay et al., 2018)



VIEWPOINT

Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness

Jacqueline Corrigan-Curay, JD, MD
Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland.

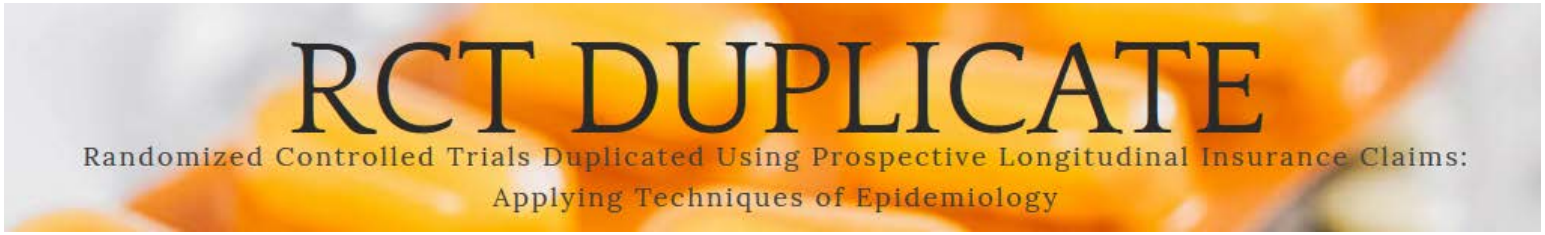
For hundreds of years, the development of new medical treatments relied on "real-world" experience. Discoveries such as citrus fruit curing scurvy described in the 1700s or insulin as a treatment for diabetes in the 1920s long preceded the advent of the modern randomized clinical trial. What these diseases had in common was a

records (EHRs), together with rising costs and recognized limitations of traditional trials, has renewed interest in the use of real-world data (RWD) to enhance the efficiency of research and bridge the evidentiary gap between clinical research and practice. RWD can be defined as data relating to patient health status or the

Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA. 2018 Sep 4;320(9):867-868.

Design: Target trial emulation

German Diabetes Center



Our Work

We are building an empirical evidence base for real world data through large-scale replication of randomized controlled trials. Our goal is to understand for what types of clinical questions real world data analyses can be conducted with confidence and with which designs and analysis methods.

Our Impact

If principled nonrandomized study approaches based on healthcare databases can consistently match the results of published trials and predict the results of ongoing trials, then we gain confidence in the validity of future real world data analyses that may be performed in the absence of randomized trial evidence.

www.rctduplicate.org

Design: Target trial emulation

DDZ German Diabetes Center

RCT DUPLICATE

Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology

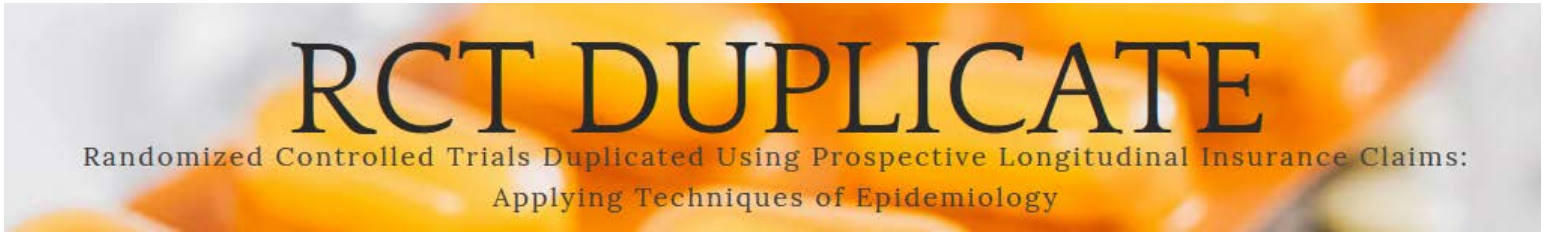
Jessica Franklin, PhD
Co-Director, RCT DUPLICATE
Assistant Professor

Sebastian Schneeweiss, MD, ScD
Co-Director, RCT DUPLICATE
Vice Chief

www.rctduplicate.org

Design: Target trial emulation

DDZ German Diabetes Center



Expert Advisory Panel

Alan Brookhart, PhD
Professor of Epidemiology and Biostatistics, Center of Pharmacoepidemiology, University of North Carolina Chapel Hill

Steven Goodman, MD, PhD
Associate Dean of Clinical and Translational Research
Professor of Medicine and of Health Research & Policy, Stanford University

Miguel Hernan, MD, DrPH
Professor of Epidemiology and Biostatistics, Harvard School of Public Health

Wayne Ray, PhD
Professor of Health Policy, Vanderbilt University School of Medicine

Samy Suissa, PhD
Director of the Centre for Clinical Epidemiology, Lady Davis Institute
Professor of Epidemiology and Biostatistics, McGill University

www.rctduplicate.org

Weitere Methoden, um für unbekanntes Confounding zu adjustieren

DDZ German Diabetes Center

- Negative Controls (Lipsitch et al., 2010)
- Prior event rate ratio (Tannen et al., 2009)

Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010 May;21(3):383-8.
Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ. 2009 Jan 27;338:b81.

Fazit

DDZ German Diabetes Center

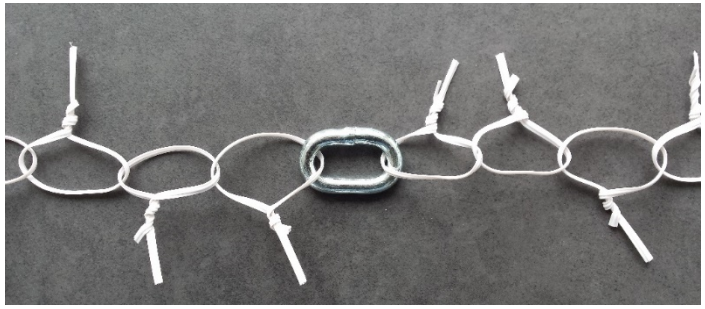
- Die randomisierte kontrollierte Studie ist nach wie vor das Design der Wahl, um die Wirksamkeit von Therapien zu prüfen.
- Die klinische Forschung sollte jedoch darauf achten, dass diese Erkenntnis nicht zum Dogma erstarrt, sondern sich auf die bestehende (und kommende) empirische Evidenz verlassen, die RCTs und Non-RCTs systematisch und mit validen Designs vergleicht.

Fazit

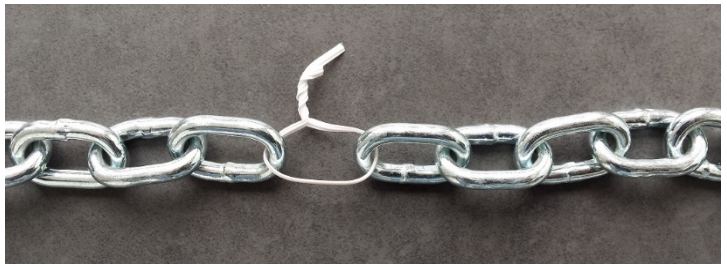
DDZ German Diabetes Center

- Das (vorläufige) Ergebnis dieses Vergleichs ist, dass **gut gemachte** Non-RCTs zur Beurteilung von Interventionen geeignet sind, v.a. vor dem Hintergrund der Schwächen, die RCTs naturgemäß haben.

RCT

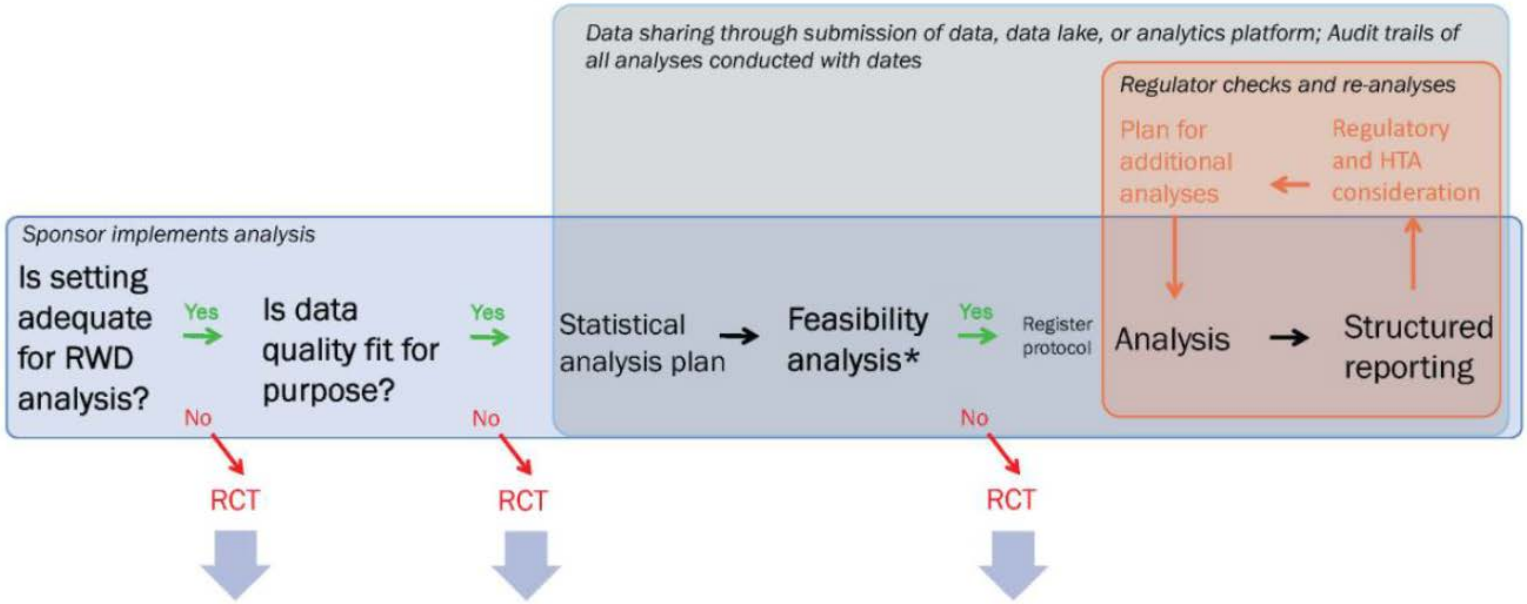


Non-RCT



A structured process for implementation of RWD analyses for regulatory decision-making

German Diabetes Center



* Feasibility analysis can include checking covariate balance after applying the chosen confounding adjustment strategy, checking statistical power, evaluating positive or negative control outcomes, and other analyses without evaluating the study outcomes in the two treatment groups.

Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. Clin Pharmacol Ther. 2019 Apr;105(4):867-877.

Fazit

DDZ German Diabetes Center

- Die Nachfrage von Patienten, Klinikern und dem gesamten Gesundheitssystem nach Evidenz aus Non-RCTs wird in den nächsten Jahren noch ansteigen. Es gibt schlichtweg zu viele Fragestellungen in der medizinischen Versorgung, als dass alle in RCTs beantwortet werden könnten.

  Zudem wird sich die Gesellschaft weder die dazu nötigen Mittel noch die dazu nötige Zeit leisten können oder wollen. (Borah et al., 2014)

Borah BJ, Moriarty JP, Crown WH, Doshi JA: Applications of propensity score methods in observational comparative effectiveness and safety research: where have we come and where should we go? J Comp Eff Res 2014; 3: 63–78.

DDZ
German Diabetes Center

Vielen Dank für Ihre Aufmerksamkeit!

Leibniz Gemeinschaft
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF
Bundesministerium für Gesundheit
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen
Bundesministerium für Bildung und Forschung
DZD Deutsches Zentrum für Diabetesforschung