Randomization and Causal Inference in Clinical Studies

Martin Schumacher, Claudia Schmoor, Jan Beyersmann

Institute for Medical Biometry and Statistics and Clinical Trials Unit
Medical Center and Faculty of Medicine – University of Freiburg, Germany
Institute for Statistics, University of Ulm, Germany

IQWIG im Dialog 2019
Köln, 21.06.2019

Contents

- The potential outcome scenario: Randomization and propensity score revisited
- A comprehensive cohort study
- Systematic reviews on comparisons of randomized and observational studies
- The EMPA-REG OUTCOME trial: From standard statistical analysis to mediation analyses
- Discussion and conclusions

Conflict of interest statement

The Institute for Medical Biometry and Statistics and the Clinical Trials Unit of the Medical Center and Faculty of Medicine – University of Freiburg, Germany obtained an institutional research grant from Boehringer Ingelheim for an independent analysis of the EMPA-REG OUTCOME trial and subsequent specific analyses requested by the Steering Committee of the study.

The potential outcome scenario (1)

- $(Y_{(0)}, Y_{(1)})$ potential outcome vector for a patient
  with $Y_{(0)}$: outcome if the control (or no) treatment is given
  $Y_{(1)}$: outcome if the new treatment is given
- Interest is in $E(Y_{(1)} - Y_{(0)})$,
  called the average causal effect, or any suitable functional
  of the joint distribution F_{01} of $(Y_{(0)}, Y_{(1)})$
- Usually (except for a perfect cross-over study), only

$$Y = X\,Y_{(1)} + (1-X)\,Y_{(0)}$$

is observed with X = 1 {new treatment is given}.

The potential outcome scenario (2)

- With randomized treatment allocation, we can identify the marginal distributions F_0 of $Y_{(0)}$ and F_1 of $Y_{(1)}$ and estimate them in an unbiased way.
- We can therefore identify and estimate the average causal effect

  $$E(Y_{(1)} - Y_{(0)}) = E(Y_{(1)}) - E(Y_{(0)})$$

  or any suitable functional of the marginal distributions F_0 and F_1 in a randomized clinical trial
- Randomization ensures balance of all known and unknown potential confounders (except for random imbalances)

The potential outcome scenario (3)

- The propensity score (PS) is defined as P (X=1|C) where C is a vector of covariates
- We can identify the average causal effect under the assumption

  $(Y_{(0)}, Y_{(1)})$ independent of X | C

  („No unmeasured confounders“)
- The assumption of „No unmeasured confounders“ implies

  $(Y_{(0)}, Y_{(1)})$ independent of X | P(X=1|C)
- Additional assumption: $0 < P(X=1|C) < 1$,
  i.e. every patient can receive either treatment

Propensity score in practice

- There are various ways using the propensity score (Matching, Weighting, Stratification, Covariate in outcome regression model)

Propensity score in practice

- There are various ways using the propensity score (Matching, Weighting, Stratification, Covariate in outcome regression model)
- The propensity score has to be estimated: how to model?
  - Commonly used: logistic regression model
  - Which covariates to include?
  - Sparse or high-dimensional model?
  - Penalized regression (e.g. lasso-type)?
  - Penalized spline imputation method?

Contents

- The potential outcome scenario: Randomization and propensity score revisited
- A comprehensive cohort study
- Systematic reviews on comparisons of randomized and observational studies
- The EMPA-REG OUTCOME trial: From standard statistical analysis to mediation analyses
- Discussion and conclusions

A comprehensive cohort study (1)

- Study conducted by the German Breast Cancer Study Group to compare three cycles of chemotherapy (3 CMF) with six cycles of chemotherapy (6 CMF) in patients with non-metastatic node-positive breast cancer
- Randomized as well as patients not consenting to randomization were enrolled and followed according to a standard protocol
- Primary endpoint: event-free survival

American Journal of Epidemiology

Vol. 167, No. 9
DOI: 10.1093/aje/kwn010
Advance Access publication March 11, 2008

Practice of Epidemiology

Evidence from Nonrandomized Studies: A Case Study on the Estimation of Causal Effects

Claudia Schmoor[1], Angelika Caputo[2], and Martin Schumacher[2]

[1] Clinical Trials Center, University Medical Center Freiburg, Freiburg, Germany.
[2] Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany.

Received for publication July 13, 2007; accepted for publication January 11, 2008.

Although randomized controlled trials are regarded as the gold standard for comparison of treatments, evidence from observational studies is still relevant. To cope with the problem of possible confounding in these studies, investigators need methods for analyzing their results which adjust for confounders and lead to unbiased estimation of the treatment effect. In this paper, the authors describe the main principles of three statistical methods for doing this. The first method is the classical approach of a multiple regression model including the effects of treatment and covariates. This considers the relation between prognostic factors and the outcome variable as a relevant criterion for adjustment. The second method is based on the propensity score, focusing on the relation between prognostic factors and treatment assignment. The third method is an ecologic approach using a grouped treatment variable, which may aid in avoiding confounding by indication. These approaches are applied to a partially randomized trial conducted in 720 German breast cancer patients between 1984 and 1997. The study had a comprehensive cohort study design that included recruitment of patients who had consented to participation but not to randomization because of a preference for one of the treatments. This design offers a unique opportunity to contrast results from the nonrandomized portion of a study with those for a randomized subcohort as a reference.

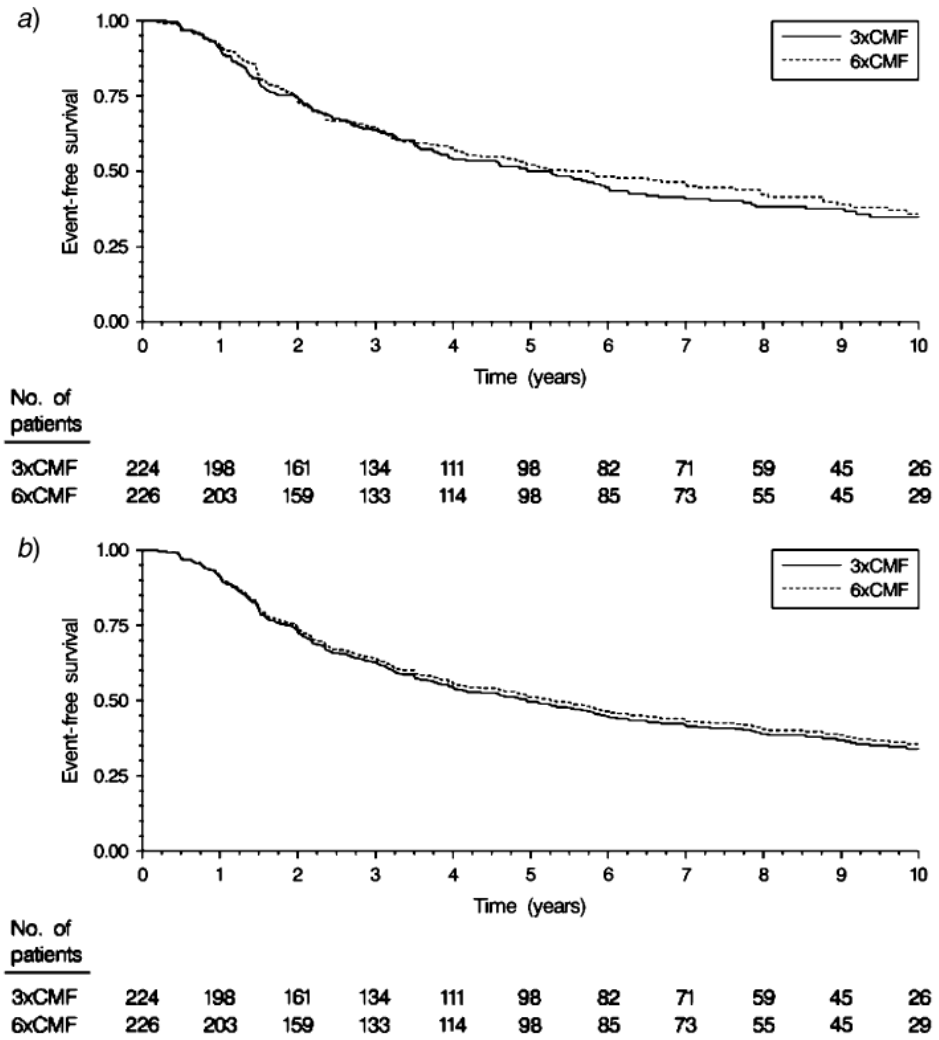
TABLE 4. Relation of covariates to individual treatment assignment and to proportion of patients treated with 3×CMF* at the respective clinical center (grouped-treatment variable) in the nonrandomized portion ($n = 238$) of a comprehensive cohort study of breast cancer, German Breast Cancer Study Group, Germany, 1984–1997

Factor	Proportion of patients treated with 3×CMF	Mean proportion of patients treated with 3×CMF at the respective clinical center (mean grouped-treatment variable)
Menopausal status		
Premenopausal	0.44	0.43
Postmenopausal	0.48	0.49
No. of positive lymph nodes		
1–3	0.60	0.53
4–9	0.33	0.38
>9	0.26	0.39
Tumor size (mm)		
≤20	0.53	0.41
21–30	0.50	0.49
>30	0.36	0.46
Tumor grade		
I	0.46	0.45
II	0.50	0.49
III	0.38	0.39
Estrogen receptor status		
Positive	0.50	0.49
Negative	0.39	0.41
Progesterone receptor status		
Positive	0.47	0.46
Negative	0.45	0.46
Treatment with tamoxifen		
No	0.40	0.45
Yes	0.61	0.49

* 3×CMF, three cycles of cyclophosphamide-methotrexate-flourouracil.

Percent nonrandomized patients receiving 3xCMF

Schmoor et al., *Am J Epidemiol* 2008



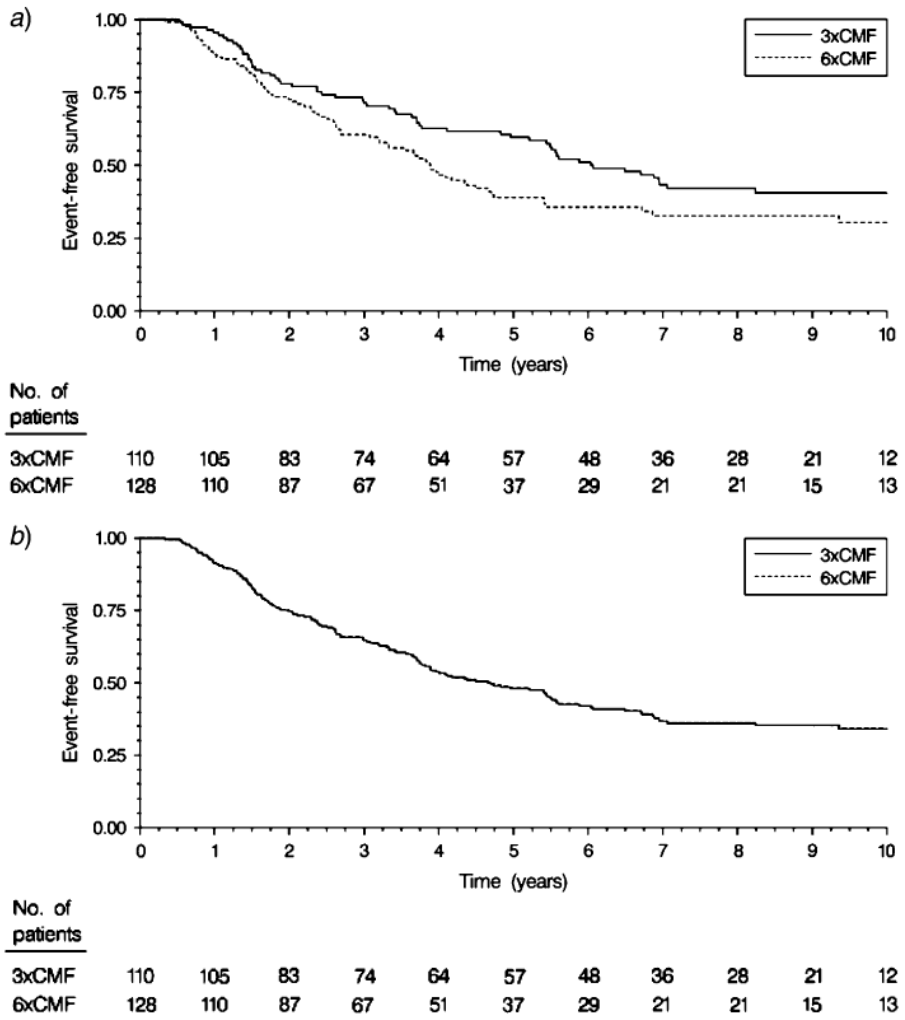
— unadjusted

Randomized patients

— adjusted

Schmoor et al., *Am J Epidemiol* 2008

FIGURE 2. Event-free survival rates by duration of chemotherapy in the randomized portion ($n = 450$) of a comprehensive cohort study of breast cancer, German Breast Cancer Study Group, Germany, 1984–1997. *a*) Unadjusted Kaplan-Meier estimates; *b*) adjusted estimates from a Cox model, adjusted for the covariates listed in table 1. 3×CMF, three cycles of cyclophosphamide-methotrexate-flourouracil; 6×CMF, six cycles of cyclophosphamide-methotrexate-flourouracil.



— unadjusted

Nonrandomized patients

— adjusted

Schmoor et al., *Am J Epidemiol* 2008

FIGURE 4. Event-free survival rates by duration of chemotherapy in the nonrandomized portion ($n = 238$) of a comprehensive cohort study of breast cancer, German Breast Cancer Study Group, Germany, 1984–1997. *a*) Unadjusted Kaplan-Meier estimates; *b*) adjusted estimates from a Cox model, adjusted for the covariates listed in table 1. In part *b*, the dotted line is not visible because the dotted line and the solid line are superimposed upon each other. 3×CMF, three cycles of cyclophosphamide-methotrexate-flourouracil; 6×CMF, six cycles of cyclophosphamide-methotrexate-flourouracil.

TABLE 3. Effect of treatment with three cycles of CMF* versus treatment with six cycles of CMF among randomized and nonrandomized breast cancer patients in unadjusted and adjusted analyses, using different methods of adjustment, German Breast Cancer Study Group, Germany, 1984–1997

Method of analysis	Hazard ratio	Standard error	95% confidence interval	*p* value†
Randomized patients ($n = 450$; 262 events)				
Unadjusted‡	1.077	0.124	0.845, 1.372	0.55
Conventional adjustment for covariates§	1.054	0.125	0.825, 1.345	0.67
Nonrandomized patients ($n = 238$; 138 events)				
Unadjusted‡	0.693	0.173	0.494, 0.973	0.034
Conventional adjustment for covariates§	1.002	0.195	0.683, 1.470	0.99
Stratified for propensity score¶	0.987	0.192	0.677, 1.438	0.95
Grouped-treatment variable#	0.758	0.280	0.438, 1.311	0.32

Schmoor et al.,
Am J Epidemiol 2008

TABLE 5. Effects of prognostic factors on treatment assignment (three cycles of CMF* vs. six cycles of CMF) in the nonrandomized portion ($n = 238$) of a comprehensive cohort study of breast cancer, German Breast Cancer Study Group, Germany, 1984–1997

Factor	Odds ratio	95% confidence interval	*p* value†
No. of positive lymph nodes			
1–3	1		<0.0001
4–9	0.29	0.16, 0.55	
>9	0.23	0.10, 0.51	
Treatment with tamoxifen			
No	1		0.003
Yes	2.54	1.38, 4.68	

* CMF, cyclophosphamide-methotrexate-flourouracil.
† *p* value from a two-sided Wald test in a logistic regression model.

Schmoor et al.,
Am J Epidemiol 2008

A comprehensive cohort study (2)

- In this particular study, propensity score as well as regression adjustment led to results very similar to those of the randomized part
- Comprehensive cohort studies have been carried out very rarely. When only the results of an observational study are available (analyzed based on a propensity score), how reliable are the results?
- Systematic comparisons of treatment effects in randomized vs. non-randomized studies?
- What are they about and what can we learn from them?

Contents

- The potential outcome scenario: Randomization and propensity score revisited
- A comprehensive cohort study
- Systematic reviews on comparisons of randomized and observational studies
- The EMPA-REG OUTCOME trial: From standard statistical analysis to mediation analyses
- Discussion and conclusions

Specific Comparisons

Reference	Medical field	Included study sample	Number of studies	Methodology used in observational studies	Direction of bias
Kuss et al. 2011 [4]	Cardiac surgery	Randomized and non-randomized studies comparing off- and on pump surgery	28 non-randomized studies and 51 randomized trials	Propensity score based analyses	Similar effects
Lonjon et al. 2014 [5]	Surgical procedures	Randomized and non-randomized studies on surgical procedures	70 non-randomized studies and 94 randomized trials	Propensity score based analyses	Similar effects
Zhang et al. 2014 [8]	Intensive Care Medicine	Randomized and non-randomized studies on treatment of patients with sepsis	14 non-randomized studies, 3 systematic reviews and 7 randomized trials	Propensity score based analyses	Overestimation of effects
Ankarfeldt et al. 2017 [2]	Diabetes	Randomized and non-randomized studies on treatment with glucose-lowering drugs	2 comparisons with 11/16 randomized studies and 7/4 non-randomized studies, published 2000-2015	Diverse	No efficacy – effectiveness gap observed

General Comparisons

Reference	Medical field	Included study sample	Number of studies	Methodology used in observational studies	Direction of bias
Kunz & Oxman, 1998 [3]	Not restricted to specific medical specialties	Cohorts or meta-analysis of clinical trials that included an empirical assessment of the relation between randomization and estimates of effects	11 comparisons with different numbers of studies published until 1998	Diverse	Over-, underestimation, reversal of effect; similar effects, “unpredictability paradox”
Odgaard-Jensen et al. 2011 [6]	Not restricted to specific medical specialties	Cohorts of studies, systematic reviews and meta-analyses of healthcare intervention that compared random vs non-random allocation	10 comparisons with different numbers of studies published until 2009	Diverse	Over- and underestimation as well as similar effects, “inconclusive results”
Anglemyer et al. 2014 [1]	Not restricted to specific medical specialties	Systematic reviews to compare effects of interventions tested in trials with those tested in observational studies	15 systematic reviews	Diverse – one comparison for propensity score based analyses	Some over- and underestimation of effects, mostly similar effects
Soni et al. 2019 [7]	Oncology	Observational studies comparing two treatment regimes for any diagnosis of cancer and matching randomized trials	350 treatment comparisons (non-randomized) and 121 randomized trials (published 2000-2016)	Diverse – “advanced statistical methods” considered	No agreement beyond what is expected by chance

original report

Comparison of Population-Based Observational Studies With Randomized Trials in Oncology

Payal D. Soni, MD[1]; Holly E. Hartman, MS[2]; Robert T. Dess, MD[2]; Ahmed Abugharib, MD[3]; Steven G. Allen, PhD[2]; Felix Y. Feng, MD[4]; Anthony L. Zietman, MD[5]; Reshma Jagsi, MD, DPhil[2]; Matthew J. Schipper, PhD[2]; and Daniel E. Spratt, MD[2]

abstract

PURPOSE Comparative efficacy research performed using population registries can be subject to significant bias. There is an absence of objective data demonstrating factors that can sufficiently reduce bias and provide accurate results.

METHODS MEDLINE was searched from January 2000 to October 2016 for observational studies comparing two treatment regimens for any diagnosis of cancer, using SEER, SEER-Medicare, or the National Cancer Database. Reporting quality and statistical methods were assessed using components of the STROBE criteria. Randomized trials comparing the same treatment regimens were identified. Primary outcome was correlation between survival hazard ratio (HR) estimates provided by the observational studies and randomized trials. Secondary outcomes included agreement between matched pairs and predictors of agreement.

J Clin Oncol 2019.

TABLE 1. Characteristics of Observational Studies (continued)

Characteristic	No. (%) of Observational Studies: All (n = 755)	Matched (n = 350)	*P*
Reporting quality			
Any age metric reported	567 (75)	267 (76)	.78
Median follow-up reported	290 (38)	138 (40)	.80
Extent of missing data reported	548 (73)	225 (71)	.54
Handling of missing data reported	475 (63)	204 (64)	.80
Statistical rigor			
Adjustments			
Age	652 (86)	312 (89)	.23
Extent of disease	654 (87)	318 (91)	.06
Comorbidities	256 (34)	124 (35)	.67
Geographic region	259 (34)	135 (39)	.19
Advanced statistical methods			
Multivariable analysis	611 (81)	292 (83)	.36
Propensity adjustment	202 (27)	124 (35)	.004
Instrumental variable	27 (4)	16 (5)	.53
Sensitivity analysis	87 (12)	50 (14)	.23

Soni et al.
J Clin Oncol 2019

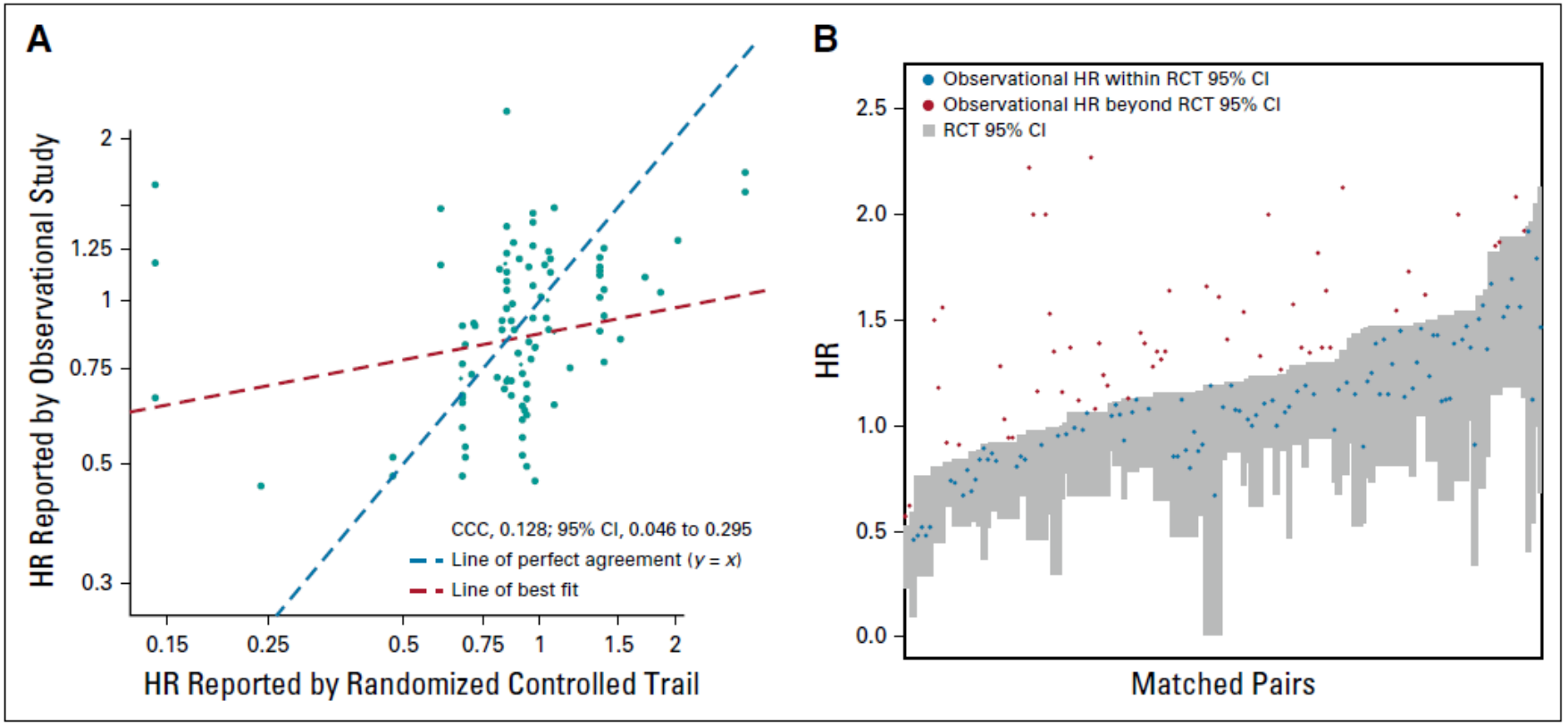


FIG 2. Comparison of hazard ratios (HRs) reported by rigorously performed, well-matched observational studies and randomized trials. (A) Scatter plot of HR reported by observational study versus randomized controlled trial (RCT) for each matched pair (n = 121). *x*- and *y*-axes presented on log scale. Red dashed line represents the line of best fit; teal dashed line represents where the line of best fit would be if the HRs from the observational study and RCT were equal. (B) RCT HR 95% CI (gray boxes) with observational study HR estimates (red and blue dots). Matched pairs ordered by the upper CI limit of RCT. HRs were inverted as necessary to ensure that both HRs were reported relative to the same reference treatment and that the observational HR was greater than the randomized trial HR. HR < 1 indicates improved survival with the comparator treatment compared with the reference. CCC, concordance correlation coefficient.

Soni et al.
J Clin Oncol 2019

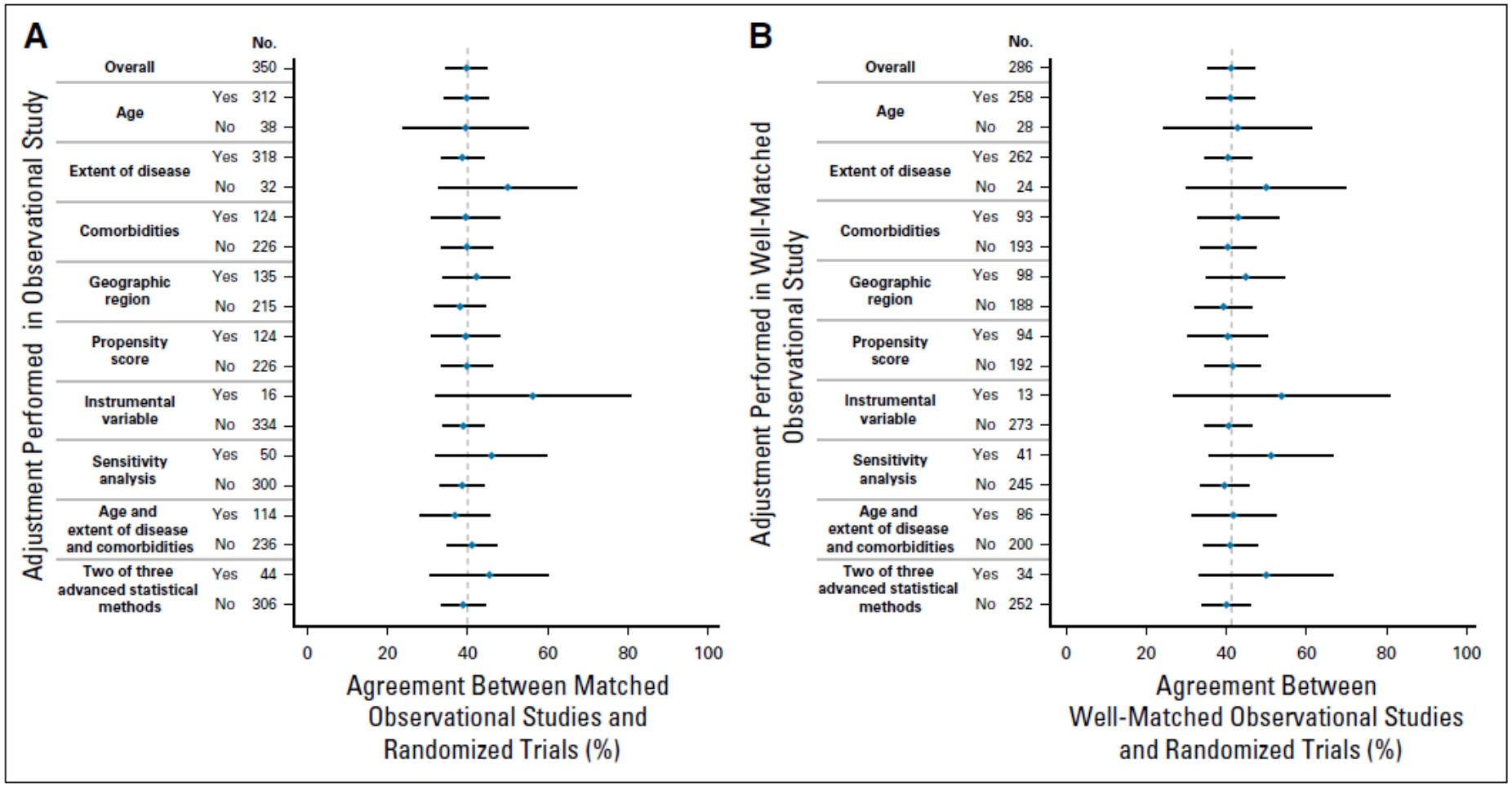


FIG 3. Agreement between observational studies and randomized controlled trials by adjustments performed in the observational study. (A) All matched observational studies and randomized trials (n = 350). (B) Well-matched observational studies and randomized trials defined by match level of 3 to 4 (n = 196). Gray line indicates match percentage for overall group.

Soni et al.
J Clin Oncol 2019

Comparison of Population-Based Observational Studies With Randomized Trials in Oncology

Payal D. Soni, MD[1]; Holly E. Hartman, MS[2]; Robert T. Dess, MD[2]; Ahmed Abugharib, MD[3]; Steven G. Allen, PhD[2]; Felix Y. Feng, MD[4]; Anthony L. Zietman, MD[5]; Reshma Jagsi, MD, DPhil[2]; Matthew J. Schipper, PhD[2]; and Daniel E. Spratt, MD[2]

CONCLUSION We were unable to identify any modifiable factor present in population-based observational studies that improved agreement with randomized trials. There was no agreement beyond what is expected by chance, regardless of reporting quality or statistical rigor of the observational study. Future work is needed to identify reliable methods for conducting population-based comparative efficacy research.

J Clin Oncol 37:1209-1216.

Effectiveness in the Absence of Efficacy: Cautionary Tales From Real-World Evidence

Safiya Karim, MD[1] and Christopher M. Booth, MD[2]

permanent colostomy rate.[24] Is it plausible that despite the results from a large RCT with long follow-up times, this study found a 10% improvement in OS but no benefit in local control? Again in this study, the large "survival benefit" observed is more likely due to residual confounding from patient characteristics that allowed complete delivery of RT rather than the RT itself.

Conclusion

Although RWD can provide valuable insight into the benefit of treatments in the real world, there are inherent limitations to this study design. Studies of comparative effectiveness are ideally performed with a multidisciplinary team involving clinicians, epidemiologists, and biostatisticians. These studies are best suited for settings in which there is existing evidence to believe that a given treatment is efficacious (ie, to understand if efficacy translates to effectiveness). In settings where RCTs do not exist or may not be feasible, RWD can be informative; however, these studies should be interpreted with caution. Clinicians should not adopt new therapies on the basis of RWE in isolation. This is particularly true when RCTs have revealed no evidence of treatment benefit; reports of "effectiveness" in this setting are more likely artifact and may be misleading. Journal editors and clinicians should be critical of studies that report effectiveness in the absence of efficacy and should question the plausibility of such findings.

J Clin Oncol 2019.

Contents

- The potential outcome scenario: Randomization and propensity score revisited
- A comprehensive cohort study
- Systematic reviews on comparisons of randomized and observational studies
- The EMPA-REG OUTCOME trial: From standard statistical analysis to mediation analyses
- Discussion and conclusions

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes

Bernard Zinman, M.D., Christoph Wanner, M.D., John M. Lachin, Sc.D., David Fitchett, M.D., Erich Bluhmki, Ph.D., Stefan Hantel, Ph.D., Michaela Mattheus, Dipl. Biomath., Theresa Devins, Dr.P.H., Odd Erik Johansen, M.D., Ph.D., Hans J. Woerle, M.D., Uli C. Broedl, M.D., and Silvio E. Inzucchi, M.D., for the EMPA-REG OUTCOME Investigators

ABSTRACT

BACKGROUND

The effects of empagliflozin, an inhibitor of sodium–glucose cotransporter 2, in addition to standard care, on cardiovascular morbidity and mortality in patients with type 2 diabetes at high cardiovascular risk are not known.

N Engl J Med 2015.

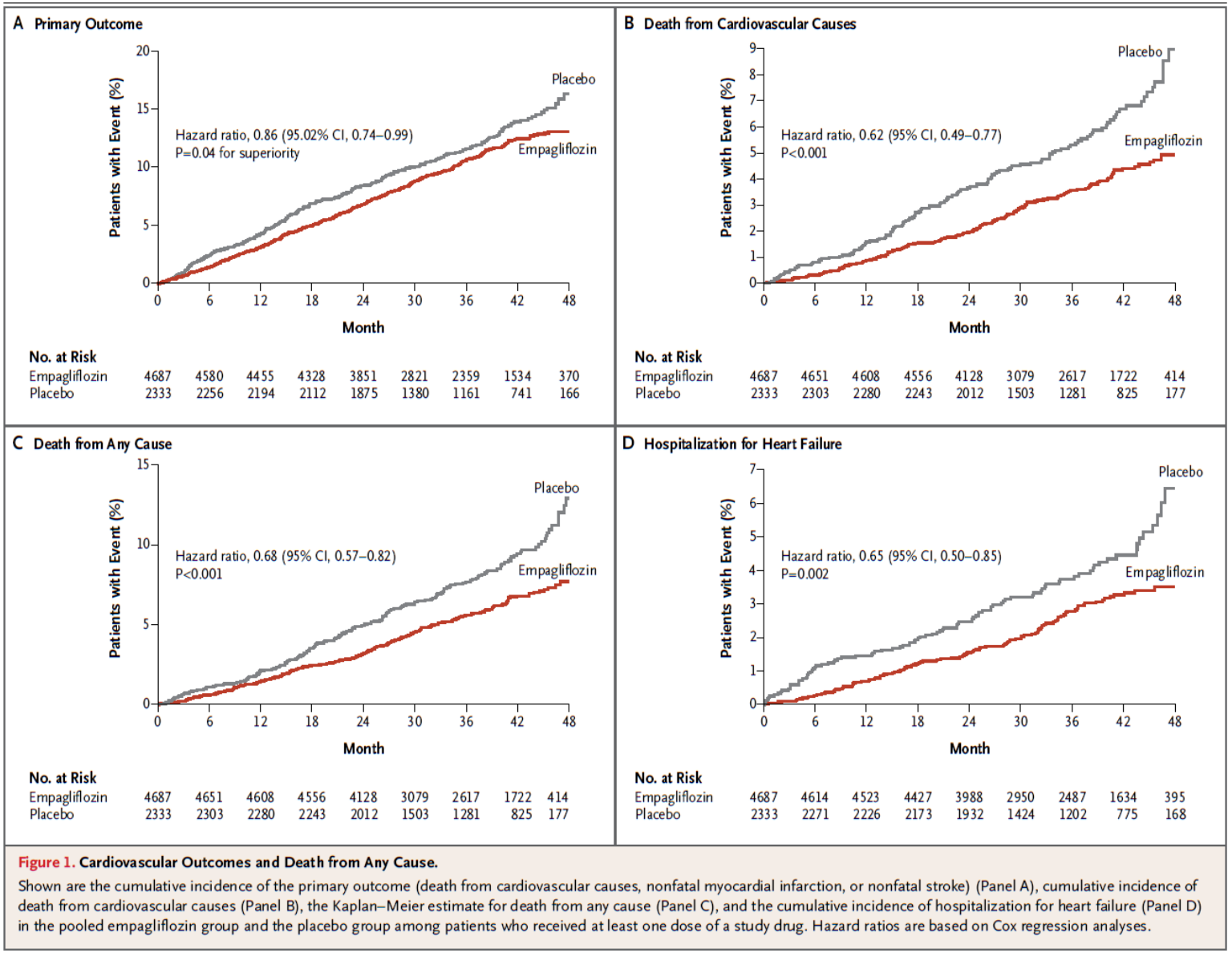


Figure 1. Cardiovascular Outcomes and Death from Any Cause.

Shown are the cumulative incidence of the primary outcome (death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke) (Panel A), cumulative incidence of death from cardiovascular causes (Panel B), the Kaplan–Meier estimate for death from any cause (Panel C), and the cumulative incidence of hospitalization for heart failure (Panel D) in the pooled empagliflozin group and the placebo group among patients who received at least one dose of a study drug. Hazard ratios are based on Cox regression analyses.

CV Death ←

Zinman et al. *NEJM,* 2015

Table 1. Primary and Secondary Cardiovascular Outcomes.

Outcome	Placebo (N=2333)		Empagliflozin (N=4687)		Hazard Ratio (95% CI)	P Value
	no. (%)	*rate/1000 patient-yr*	*no. (%)*	*rate/1000 patient-yr*		
Death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke: primary outcome*	282 (12.1)	43.9	490 (10.5)	37.4	0.86 (0.74–0.99)	
Noninferiority						<0.001†
Superiority						0.04†
Death from cardiovascular causes, nonfatal myocardial infarction, nonfatal stroke, or hospitalization for unstable angina: key secondary outcome*	333 (14.3)	52.5	599 (12.8)	46.4	0.89 (0.78–1.01)	
Noninferiority						<0.001†
Superiority						0.08†
Death						
From any cause	194 (8.3)	28.6	269 (5.7)	19.4	0.68 (0.57–0.82)	<0.001
From cardiovascular causes	137 (5.9)	20.2	172 (3.7)	12.4	0.62 (0.49–0.77)	<0.001
Fatal or nonfatal myocardial infarction excluding silent myocardial infarction	126 (5.4)	19.3	223 (4.8)	16.8	0.87 (0.70–1.09)	0.23
Nonfatal myocardial infarction excluding silent myocardial infarction	121 (5.2)	18.5	213 (4.5)	16.0	0.87 (0.70–1.09)	0.22

CV Death

Zinman et al. *NEJM,* 2015

PERSPECTIVE

Is the Mortality Benefit With Empagliflozin in Type 2 Diabetes Mellitus Too Good To Be True?

Sanjay Kaul, MD

Signaling a likely end to a long and elusive quest for cardiovascular outcome benefit associated with treatment intervention in type 2 diabetes mellitus, the results of the EMPA-REG OUTCOME trial [BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients] were received with a standing ovation at the European Association for the Study of Diabetes scientific meeting in Stockholm, Sweden, on September 17, 2015.[1] Witnessing the spontaneous applause, I had mixed emotions. Was it time to bring the trumpets out and rejoice that the "holy grail" had finally been achieved? Or, was it more appropriate to curb the enthusiasm and question the "historic milestone," given that the mortality benefit was unexpected and unprecedented?

Kaul S. *Circulation* 2016.

Cardiovascular mortality and morbidity in patients with type 2 diabetes following initiation of sodium-glucose co-transporter-2 inhibitors versus other glucose-lowering drugs (CVD-REAL Nordic): a multinational observational analysis

Kåre I Birkeland, Marit E Jørgensen, Bendix Carstensen, Frederik Persson, Hanne L Gulseth, Marcus Thuresson, Peter Fenici, David Nathanson, Thomas Nyström, Jan W Eriksson, Johan Bodegård, Anna Norhammar

Summary

Background In patients with type 2 diabetes and a high cardiovascular risk profile, the sodium-glucose co-transporter-2 (SGLT2) inhibitors empagliflozin and canagliflozin have been shown to lower cardiovascular morbidity and mortality. Using real-world data from clinical practice, we aimed to compare cardiovascular mortality and morbidity in new users of SGLT2 inhibitors versus new users of other glucose-lowering drugs, in a population with a broad cardiovascular risk profile.

Lancet Diabetes Endocrinol 2017; 15:709–17

Published Online
August 3, 2017
http://dx.doi.org/10.1016/S2213-8587(17)30258-9

See Comment page 673

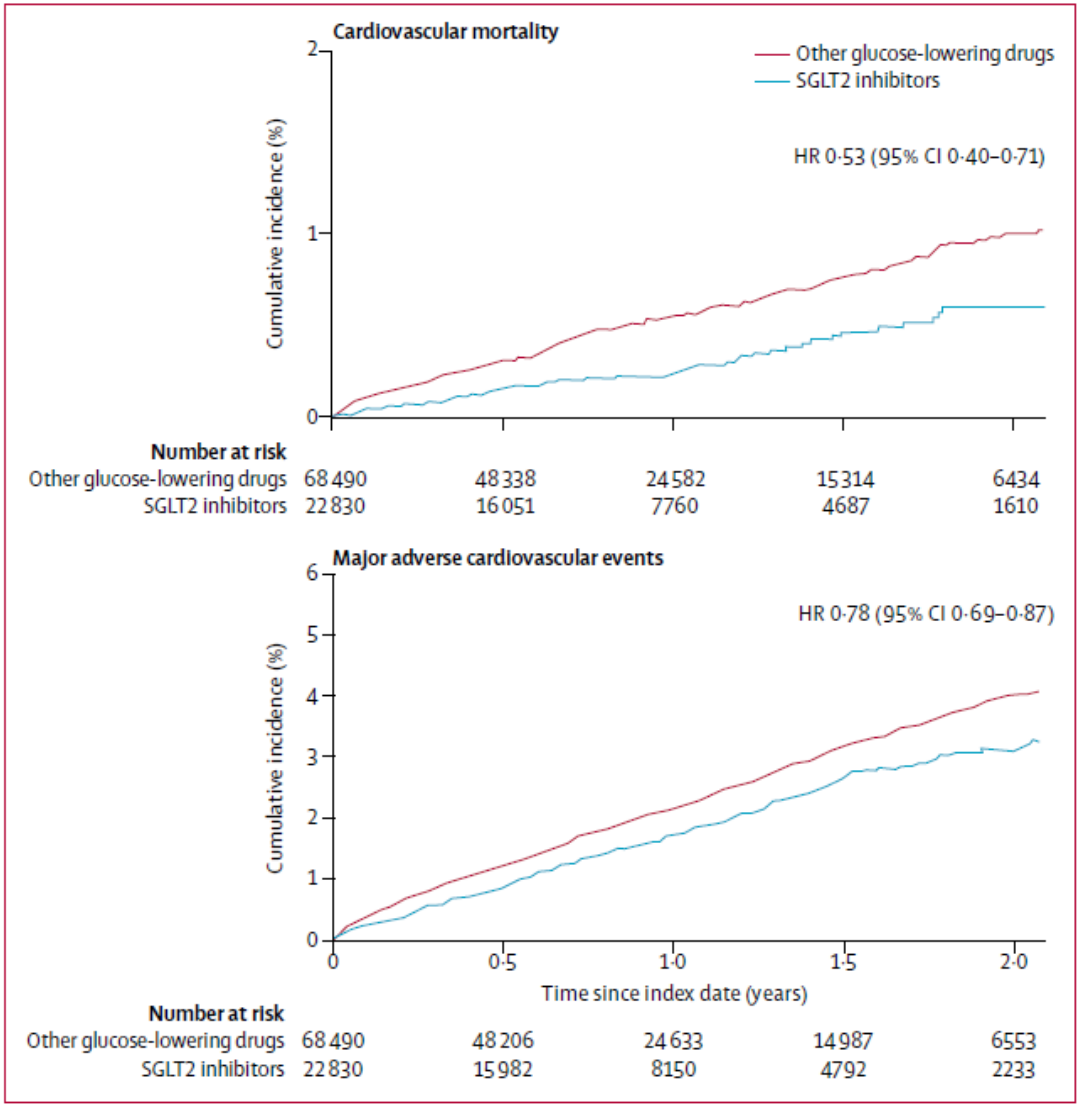


***Figure 2*: Pooled Kaplan-Meier curves and hazard ratios comparing new users of SGLT2 inhibitors and new users of other glucose-lowering drugs for cardiovascular mortality and major adverse cardiovascular events**
Groups were matched 1:3 by propensity score. SGLT2=sodium-glucose co-transporter-2. HR=hazard ratio.

← **CV Death**

Birkeland et al. *Lancet Diabetes Endocrinol* 2017

SGLT2 inhibitors in the real world: too good to be true?

	Proportion of SGLT2 inhibitor used in treated group*	Study duration, years (mean)	MACE (HR [95% CI])	Cardiovascular mortality (HR [95% CI])	All-cause mortality (HR [95% CI])	Heart failure outcome† (HR [95% CI])
CVD-REAL Nordic[2]	Dapagliflozin 94%	0·9	0·78 (0·69–0·87)	0·53 (0·40–0·71)	0·51, (0·45–0·58)	0·70 (0·61–0·81)
CVD-REAL US cohort[3]	Canagliflozin 75–76% Dapagliflozin 19%	0·5	NA	NA	0·38 (0·29–0·50)	0·55 (0·44–0·69
EMPA-REG OUTCOME[1]	Empagliflozin 100%	3·1	0·86 (0·74–0·99)	0·62 (0·49–0·77)	0·68 (0·57–0·82)	0·65 (0·50–0·85)
CANVAS[4]	Canagliflozin 100%	3·6	0·86 (0·75–0·97)	0·87 (0·72–1·06)	0·87 (0·74–1·01)	0·67 (0·32–0·87)

SGLT2=sodium-glucose co-transporter 2. MACE=major adverse cardiosvascular event (defined as cardiovascular mortality, non-fatal myocardial infarction, a ratio. NA=not available. *Defined as exposure time in propensity-matched cohort in CVD-REAL studies.[2,3] †Heart failure outcomes were defined as: hospital a and CANVAS[4]; hospital admission for heart failure (in the USA, UK, and Germany) and as inpatient or outpatient hospital visits for heart failure (Sweden, Nor outpatient hospital visits for heart failure in CVD-REAL Nordic.[2]

***Table*: Cardiovascular outcomes in CVD-REAL, EMPA-REG OUTCOME, and CANVAS**

Published Online
August 4, 2017
http://dx.doi.org/10.1016/
S2213-8587(17)30259-0

Fitchett D. *Lancet Diabetes Endocrinol* 2017.

PERSPECTIVE

Reduced Mortality With Sodium-Glucose Cotransporter–2 Inhibitors in Observational Studies

Avoiding Immortal Time Bias

Samy Suissa, PhD

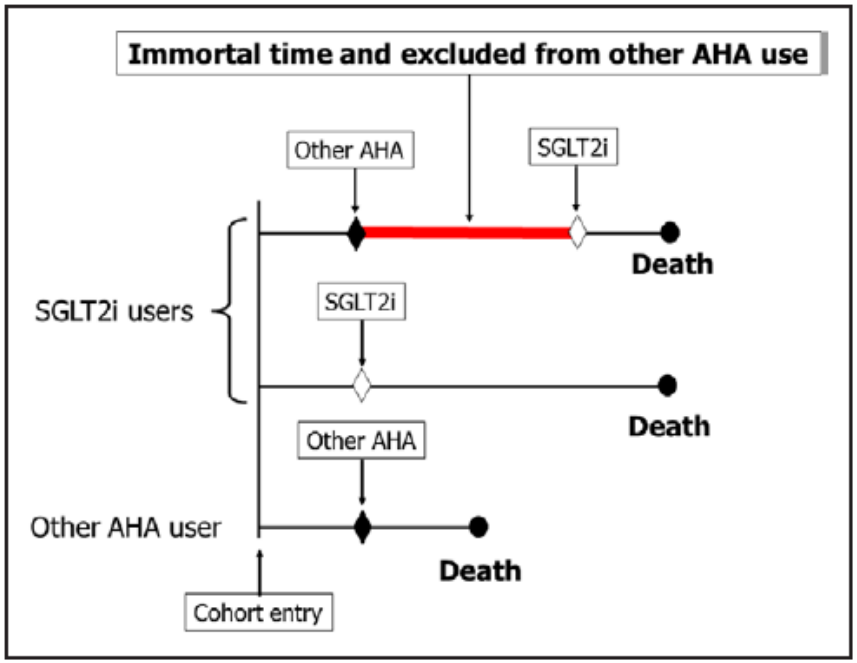


Figure. Depiction of immortal time bias: description of 2 SGLT2i-exposed and 1 SGLT2i-unexposed patients who die of any cause used in the EASEL study.[2]
The top patient initiated treatment with a non-SGLT2i drug and subsequently switched to SGLT2i, but was classified as a SGLT2i user. The time between the first non-SGLT2i AHA prescription and the first SGLT2i prescription is thus immortal (thick red line), because the subject must survive to receive this first SGLT2i prescription, but is not included as exposed to non-SGLT2i AHA, leading to immortal time bias. AHA indicates antihyperglycemic agent; and SGLT2i, sodium-glucose cotransporter–2 inhibitor.

Suissa S. *Circulation* 2018.

World Journal of Diabetes

Submit a Manuscript: https://www.f6publishing.com

World J Diabetes 2018 December 15; 9(12): 252-257

DOI: 10.4239/wjd.v9.i12.252

ISSN 1948-9358 (online)

ORIGINAL ARTICLE

Observational Study

Using real world data to assess cardiovascular outcomes of two antidiabetic treatment classes

Manfred Paul Stapff

Manfred Paul Stapff, CMO, TriNetX Inc., Cambridge, MA 02140, United States

ORCID number: Manfred Paul Stapff (0000-0002-6127-1109).

AIM

To evaluate the effect on cardiovascular outcomes of sodium-glucose co-transporter-2 (SGLT2) inhibitors in a real world setting by analyzing electronic medical records.

METHODS

We used TriNetX, a global federated research network providing statistics on electronic health records (EHR). The analytics subset contained EHR from approximately 38 Million patients in 35 Health Care Organizations in the United States. The records of 46,909 patients who had taken SGLT2 inhibitors were compared to 189,120 patients with dipeptidyl peptidase (DPP) 4 inhibitors. We identified five potential confounding factors and built respective strata: elderly, hypertension, chronic kidney disease (CKD), and co-medication with either insulin or metformin. Cardiovascular events were counted

Table 2 Results from the patient subgroups (strata) with potential confounding factors

	Stratum 1 > 60 yr		Stratum 2 hypertension		Stratum 3 CKD		Stratum 4 insulin		Stratum 5 metformin	
	SGLT2	control	SGLT2	control	SGLT2	control	SGLT2	control	SGLT2	control
n	23594	131219	27499	115703	3786	34388	24395	90978	37762	136569
patients with stroke or MI	9784 (6.3%)		10827 (7.6%)		4755 (12.5%)		8976 (7.8%)		8629 (4.9%)	
n in group	1077	8707	1452	9375	391	4364	1275	7701	1394	7235
percent in group	4.60%	6.60%	5.30%	8.10%	10.30%	12.70%	5.20%	8.50%	3.70%	5.30%
RR SGLT2 *vs* control	0.69		0.65		0.81		0.62		0.7	

SGLT2: Sodium-glucose co-transporter-2; MI: Myocardial infarction; RR: Risk ratio; CKD: Chronic kidney disease.

Stapff, *World J Diabetes* 2018.

Diabetes Care Volume 41, February 2018

Mortality Reduction in EMPA-REG OUTCOME Trial: Beyond the Antidiabetes Effect

Samy Suissa

Diabetes Care 2018;41:219–223 | https://doi.org/10.2337/dc17-1059

Two recent large-scale cardiovascular outcome trials, a now common tool in assessing the safety of pharmacological treatments for type 2 diabetes, reported significant reductions in all-cause mortality. In EMPA-REG OUTCOME [BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients], patients who received the SGLT2 inhibitor empagliflozin had a notable reduction of 9.2 deaths per 1,000 per year, while LEADER (Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results—A Long Term Evaluation) found that the patients receiving the GLP-1 receptor agonist liraglutide had a reduction of 3.7 deaths per 1,000 per year. The hypotheses to explain the sizable mortality reduction in EMPA-REG OUTCOME have mainly focused on the

How Does Empagliflozin Reduce Cardiovascular Mortality? Insights From a Mediation Analysis of the EMPA-REG OUTCOME Trial

Silvio E. Inzucchi,[1] Bernard Zinman,[2] David Fitchett,[3] Christoph Wanner,[4] Ele Ferrannini,[5] Martin Schumacher,[6] Claudia Schmoor,[6] Kristin Ohneberg,[6] Odd Erik Johansen,[7] Jyothis T. George,[8] Stefan Hantel,[9] Erich Bluhmki,[9] and John M. Lachin[10]

Diabetes Care 2018;41:356–363 | https://doi.org/10.2337/dc17-1096

OBJECTIVE

In the BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG OUTCOME) trial involving 7,020 patients with type 2 diabetes and established cardiovascular (CV) disease, empagliflozin given in addition to standard of care reduced the risk of CV death by 38% versus placebo (hazard ratio [HR] 0.62 [95% CI 0.49, 0.77]). This exploratory mediation analysis assesses the extent to which treatment group differences in covariates during the trial contributed to CV death risk reduction with empagliflozin.

[1]Section of Endocrinology, Yale University School of Medicine, New Haven, CT
[2]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, and Division of Endocrinology, University of Toronto, Toronto, Ontario, Canada
[3]St. Michael's Hospital, Division of Cardiology, University of Toronto, Toronto, Ontario, Canada

Potential mediators

MECHANISTIC CATEGORY	Variable	Name
GLYCEMIA	HbA1c	HBA1C
	Fasting Plasma Glucose	FPG
VASCULAR TONE	Systolic BP	SBP
	Diastolic BP	DBP
	Heart Rate	HR
LIPIDS	HDL-C	HDL
	LDL	LDL
	Triglycerides	TRIGL
RENAL	Urine Albumin: Cr Ratio	logUACR
	eGFR (MDRD)	EGFRM
	eGFR (CKD-num EPI)	EGFRC
BODY MASS	Weight	WEIGHT
	BMI	BMI
	Waist Circumference	WAIST
VOLUME	Hematocrit	HCT
	Hemoglobin	HGB
	Albumin	ALB
OTHER	Uric Acid	URIC

Traditional Mediation Analysis

A traditional mediation analysis as originally proposed by Baron and Kenny (12) was used, taking the time-dynamic evolvement of both the potential mediators and the outcome CV death into account. A variable must satisfy several conditions to be a mediator of the treatment effect. Treatment must have an effect on the variable over time, and the change in the variable over time must have an effect on the outcome. As an additional condition, in an analysis where the variable is included as a time-dependent covariate over time, the effect of treatment on the outcome (represented as the HR) must be reduced compared with the treatment effect in an unadjusted analysis.

Inzucchi et al.
Diabetes Care 2018.

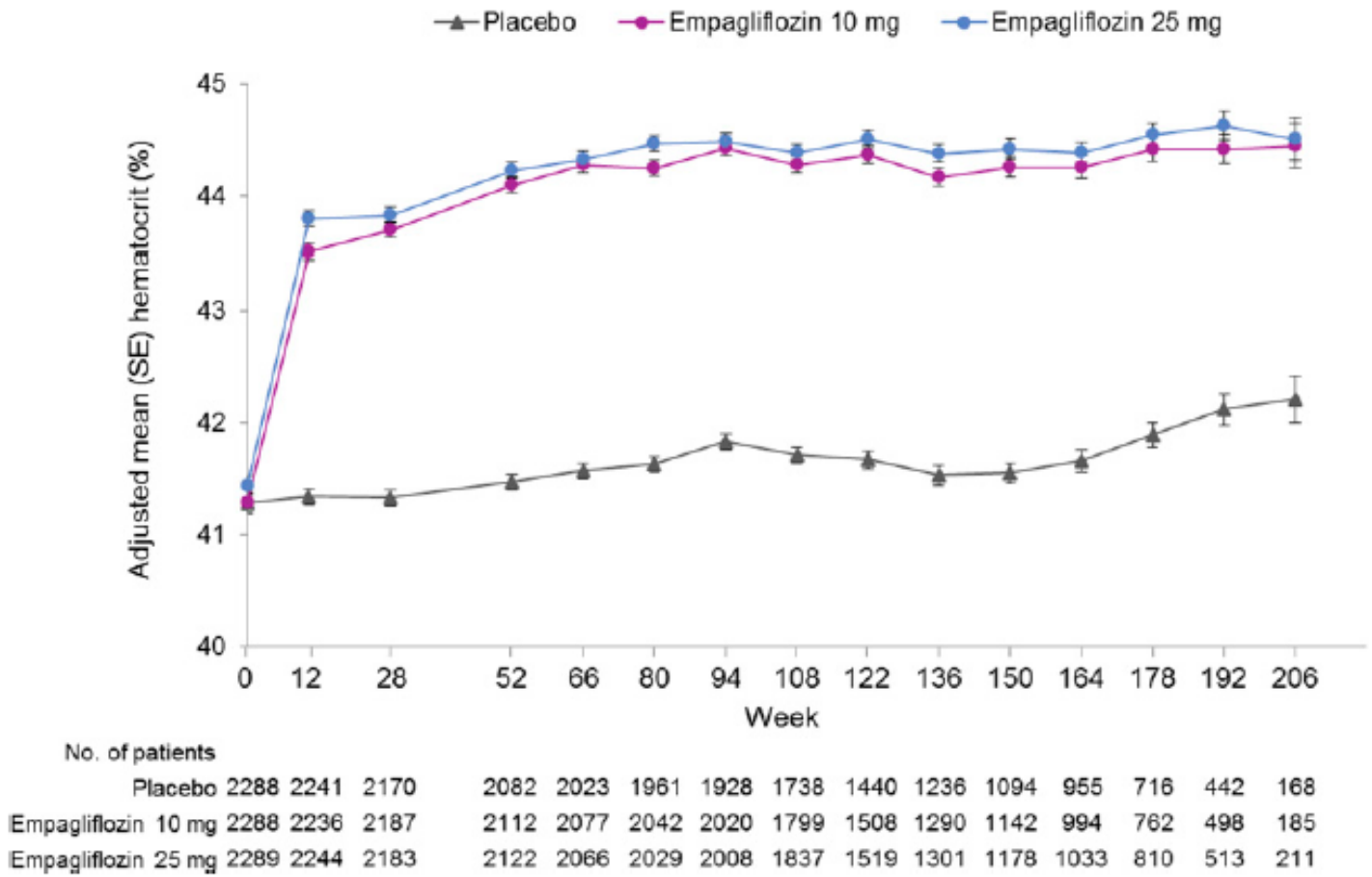


Figure 1—Hematocrit over time in patients treated with empagliflozin 10 mg, empagliflozin 25 mg, and placebo. Mixed-model repeated-measures analysis using all data up to individual trial completion in patients treated with one or more doses of study drug who had a baseline and postbaseline measurement.

Inzucchi et al. *Diabetes Care* 2018.

Model building strategy

- Starting with bivariable Cox regression models of the effect of treatment and the potential mediators M on outcome Y, one at a time separately for all potential mediators
- Multivariable Cox regression model with one representative of the different mechanistic categories
- Variable being the most promising with regard to its potential as mediator was chosen as representative
- Only variables chosen, which showed an effect on outcome Y and which led to a reduced treatment effect estimate (hazard ratio, HR, shifted to one) in the bivariable models
- For ranking of the strength of mediators:
  Multivariable model building with step-up procedure including in each step additionally the variable with the most mediating effect

SUPPLEMENTARY DATA

Supplementary Table S3. Final multivariable analysis built from step-up procedure including variables from different mechanistic categories leading to maximal mediation of treatment effect.
Effects of treatment and variables on risk of cardiovascular death (including the change from baseline in each variable as a time-dependent covariate, adjusted for the baseline value of each variable).

	HR for CV death	95% confidence interval	Percentage mediation
Effect of empagliflozin vs placebo Unadjusted	0.615	0.491, 0.770	--
adjusted for: FPG, logUACR, hematocrit, uric acid	0.931	0.732, 1.183	85.2
Effect of a 1-unit increase in:			
FPG (mg/dL)	1.003	1.001, 1.006	–
logUACR (1.0 measured on log-scale log[mg/g])	1.213	1.089, 1.351	–
Hematocrit (%)	0.919	0.892, 0.947	–
Uric acid (mg/dL)	1.291	1.186, 1.406	–

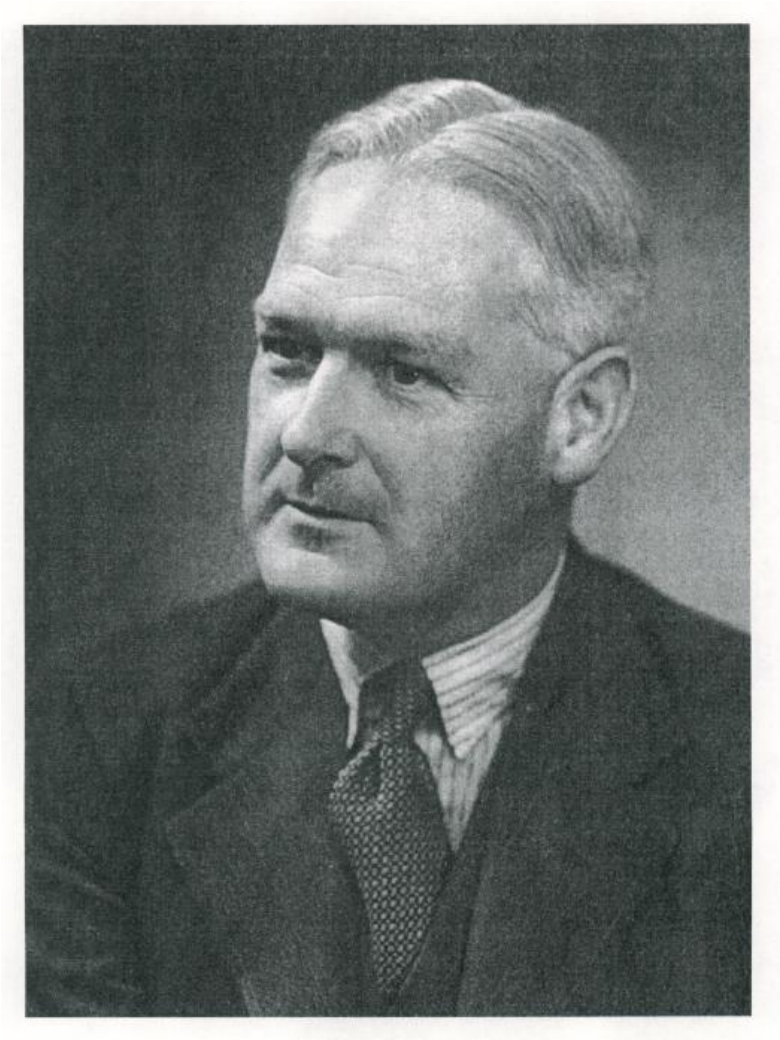
Cox regression analysis in patients treated with ≥1 dose of study drug. CV, cardiovascular; FPG, fasting plasma glucose; UACR, urine albumin:creatinine ratio.

Inzucchi et al. *Diabetes Care* 2018.

Contents

- The potential outcome scenario: Randomization and propensity score revisited
- A comprehensive cohort study
- Systematic reviews on comparisons of randomized and observational studies
- The EMPA-REG OUTCOME trial: From standard statistical analysis to mediation analyses
- Discussion and conclusions

Bradford Hill‘s criteria (1965)

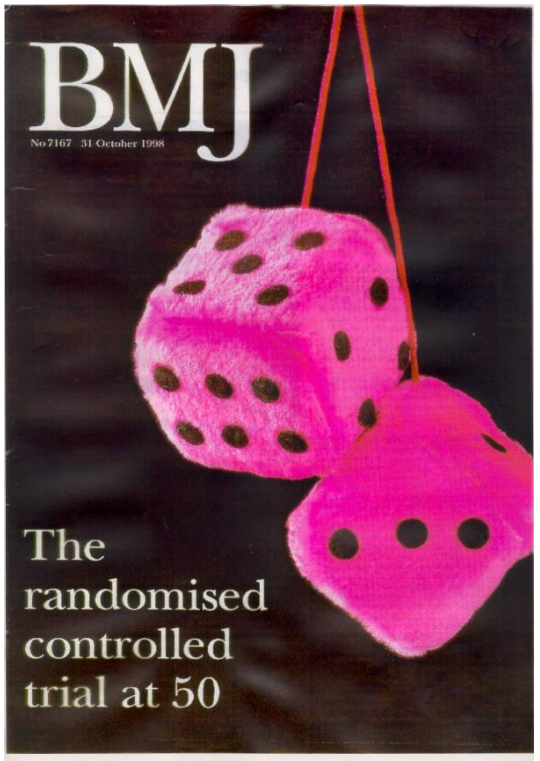


1. Strength of Association
2. Consistency
3. Specificity
4. Temporality
5. Biological Gradient
   (dose response)
6. Plausibility
7. Coherence
8. Experimental Evidence
9. Analogy

Hill AB, Proc Royal Soc Med 1965

Sir Austin Bradford Hill

- Presentation of „Principles of Medical Statistics“ (The Lancet, 1937)
- Randomized trial on streptomycin in patients with pulmonary tuberculosis (BMJ, 1948)
- Demonstration of connection between cigarette smoking and lung cancer (with Richard Doll, BMJ, 1954)



Discussion and Conclusions (1)

- Methods of causal inference, e.g. propensity score analyses, rely on assumptions that cannot be verified with the data usually available.
- Most critical is the assumption of
  "no unmeasured confounders"
  and the inclusion of confounders into a propensity score model.
- This assumption is automatically fulfilled when randomization is employed
  ("Design trumps analysis").

Discussion and Conclusions (2)

- Empirical comparisons of treatment effects in randomized trials and observational studies do not paint a clear picture. Some are themselves susceptible to bias ("A bias in the evaluation bias…", Franklin et al. *Epidemiol Methods* 2017)
- Improvement of methodology for such comparisons is urgently needed in order to not compare "apples and oranges" but "apples and apples" (Lodi et al., *Am J Epidemiol* 2019).
- Treatment effects based on observational studies are often susceptible to other sources of bias, e.g. time-related biases, besides confounding. Thus, all sources of bias have to be considered!

Discussion and Conclusions (3)

- Methods for causal inference are best suited in situations when randomization is not feasible in order to obtain the best possible evidence.
- They are also useful in randomized trials in order to address specific complications, e.g. non-compliance, treatment cross-over etc.
- As shown for the EMPAREG-Outcome trial, they can help answer additional questions on mechanisms of treatment.
- Instead of a traditional mediation analysis, more refined methods can be used (e.g. Aalen et al. *Biom J* 2019)

Take-Home Message

Randomize if you can,

Model if you must!

Modified according to J. Hanley, 2019

References – Propensity Scores

Graf E. The propensity score in the analysis of therapeutic studies. *Biom J* 1997, 39(3):297-307.

Karim S, Booth CM. Effectiveness in the absence of efficacy: Cautionary tales from real-world evidence. *J Clin Oncol* 2019; 37(13):1047-50.

Kuss O, Blettner M, Börgermann J. Propensity score: an alternative method of analyzing treatment effects. *Dtsch Arztebl Int* 2016, 113:597-603. doi 10.3238/arztebl.2016.0597.

Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Statistics in Medicine* 1996, 15: 263-271.

Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: a case study on the estimation of causal effects. *Am J Epidemiol* 2008, 167(9):1120-9. doi 10.1093/aje/kwn010.

Schmoor C, Gall C, Stampf S, Graf E. Correction of confounding bias in non-randomized studies by appropriate weighting. *Biom J* 2011, 53(2): 369-87. doi 10.1002/bmj.201000154.

Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018, 2005-2014. doi 10.1093/ije/dyy120.

Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: From naïve enthusiasm to intuitive understanding. *Stat Methods Med Res*2011, 21(3):273-293. doi 10.1177/0962280210394483.

Zhou T, Elliott MR, Little RJA. Penalized spline of propensity methods for treatment comparison. *J Am Stat Assoc* 2019, 114(525):1-38. doi 10.1080/01621459.2018.1518234.

References – Empirical Comparisons

[1] Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database of Systematic Reviews* 2014, Issue 4.

[2] Ankarfeldt MZ, Adalsteinsson E, Groenwold RHH, Ali MS, Klungel OH. A systematic literature review on the efficacy-effectiveness gap: comparison of randomized controlled trials and observational studies of glucose-lowering drugs. *Clin Epidemiol* 2017; 9: 41-51.

[3] Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317:1185-90.

[4] Kuss O, Legler T, Börgemann J. Treatment effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *J Clin Epidemiol* 2011; 64:1076-84.

[5] Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, Ravaud P. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg* 2014; 259: 18-25.

[6] Odgaard-Jensen J, Vist GE, Timmer A,Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2011, Issue 4.

[7] Soni PD, Hartman HE, Dess RT, Abugharib A, Allen SG, Feng FY, Zietman AL, Jagsi R, Schipper MJ, Spratt DE. Comparison of population-based observational studies with randomized trials in oncology. *J Clin Oncol* 2019; 37:1209-1216.

[8] Zhang Z, Ni H, Xu X. Do the observational studies using propensity score analysis agree with randomized controlled trials in the area of sepsis? *J Crit Care* 2014; 29:886e9-886e15.

References – EMPAREG Outcome Study

Birkeland KI, Jørgensen ME, Carstensen B, Persson F, Gulseth HL, Thuresson M, Fenici P, Nathanson D, Nyström T, Eriksson JW, Bodegård J, Norhammar A. Cardiovascular mortality and morbidity in patients with type 2 diabetes following initiation of sodium-glucose co-transporter-2 inhibitors versus other glucoose-lowering drugs (CVD-REAL Nordic): a multinational observational analysis. *Lancet Diabetes Endocrinol* 2017; 15:709-17. doi: 10.1016/S2213-8587(17)30258-9.

Fitchett D. SGLT2 inhibitors in the real world: too god to be true? *Lancet Diabetes Endocrinol* 2017; 59: 673-675.

Inzucchi SE, Zinman B, Fitchett D, Wanner C, Ferrannini E, Schumacher M, Schmoor C, Ohneberg K, Johansen OE, George JT, Hantel S, Bluhmki E, Lachin JM. How does empagliflozin reduce cardiovascular mortality? Insights from a mediation analysis of the EMPA-REG OUTCOME Trial. *Diabetes Care* 2018; 41(2) :356-363. doi: 10.2337/dc17-1096.

Kaul S. Is the mortality benefit with empagliflozin in type 2 diabetes mellitus too good to be true? *Circulation* 2016; 134: 94-96.

Stapff MP. Using real world data to assess cardiovascular outcomes of two antidiabetic treatment classes. *World J Diabetes* 2018;9(12):252-257. doi: 10.4239/wjd.v9.i12.252.

Suissa S. Mortality reduction in EMPA-REG OUTCOME trial: Beyond the antidiabetes effect. *Diabetes Care* 2018; 41:219-223.

Suissa S. Reduced mortality with sodium-glucose cotransporter-2 inhibitors in observational studies avoiding immortal time bias. *Circulation* 2018; 137(14) :1432-1434. doi: 10.1161/CIRCULATIONAHA.117.032799.

Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2015, 373(22): 2117-28.

Further References

Aalen OO, Stensrud MJ, Didelez V, Daniel R, Røysland K, Strohmaier S. Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. Biom J 2019;1-8.

Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. J Clin Epidemiol 2018; 96:12-22.

Franklin JM, Dejene S, Huybrechts KF, Wang SV, Kulldorff M, Rothman KJ. A Bias in the evaluation of bias comparing randomized trials with nonexperimental studies. Epidemiol Methods 2017; 6(1)

Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, Babiker A, Law M, Chu H, Byrne D, Horban A, Sterne J, Porter K, Sabin CA, Costagliola D, Abgrall S, Gill M, Touloumi G, Pacheco AG, van Sighem A, Reiss P, Bucher HC, Giménez A, Jarrin I, Wittkop L, Meyer L, Pérez-Hoyos S, Justice A, Neaton JD, Hernán MA. Effect estimates in randomized trials and observational studies: comparing apples with apples. Am J Epidemiol 2019.