Standardized Assessment Tool Designed to Assist in the Evaluation of RWE on Drug Effectiveness and Safety

ISPOR North America - Boston
May 9, 2023
10:15 am - 11:15 am

Our session

DISCUSSION LEADER

Ashley Jaksa, MPH
Market Access Scientific Strategy Lead
Aetion

DISCUSSANTS

Katsiaryna Bykov, PharmD, ScD
Assistant Professor of Medicine,
Harvard Medical School

Jessica Franklin, PhD
Principal Consultant, Epidemiology and RWE
Optum

Cynthia Girman, DrPH, FISPE
President
CERobs Consulting

Disclosures

Ashley Jaksa is an employee at Aetion Inc.

Katsiaryna Bykov – funded by NIH and AHRQ; consultant for Alosa Health (nonprofit)

Jessica Franklin is an employee at Optum Epidemiology

Cynthia Girman – Founder/ & President of CERobs Consulting; Methodology Committee of PCORI

The creation of the questionnaire that will be discussed today is funded in part by International Society for Pharmacoepidemiology (ISPE).

ispe International Society for Pharmacoepidemiology

Acknowledgements

Writing team:

Katsiaryna Bykov and Elisabetta Patorno – Harvard Medical School, USA

Kinwei Chan – National Taiwan University Health Data Research Center, Taiwan

Robert Epstein – Epstein Health, LLC, USA

Madlen Gazarian – University of New South Wales, Sydney, Australia

Jessica Franklin – Optum Epidemiology, USA

Cynthia Girman – CERobs Consulting, LLC, University of North Carolina (adjunct faculty), USA

Ashley Jaksa – Aetion, USA

Stephen Duffield– National Institute for Health and Care Excellence (NICE), UK

Jennifer Lund – University of North Carolina, USA

Hongbo Yuan - Canadian Agency for Drugs and Technologies in Health (CADTH), Canada

Agenda

1. Overview of the project
2. Why do we need “another” tool to evaluate bias?
3. Overview of the paper we will use to demonstrate the questionnaire
4. Evaluating the risk of bias using the questionnaire - in depth review of each domain
    - Study design
    - Misclassification bias
    - Study design biases
    - Bias due to confounding

    Katsiaryna Bykov

    - Propensity scores
    - Missing data

    Jessica Franklin

    - Summary

    Cynthia Girman
5. Q&A

Objective of the working group

To create a comprehensive, fit-for-purpose, and credible tool to streamline and harmonize RWE evaluation for HTA agencies

- Useful for non-pharmacoepidemiologists
- Tailored to HTAs (but can be used in other settings)
- Cover most sources of potential bias in RWE studies of medications
- Provide consistent and comprehensive evaluation of RWE rigor

The focus is on comparative medication safety and effectiveness studies and the validity of these studies

One barrier to using RWE is lack of expertise in observational study design and methods

Unfamiliarity and lack of knowledge on RWE methodology

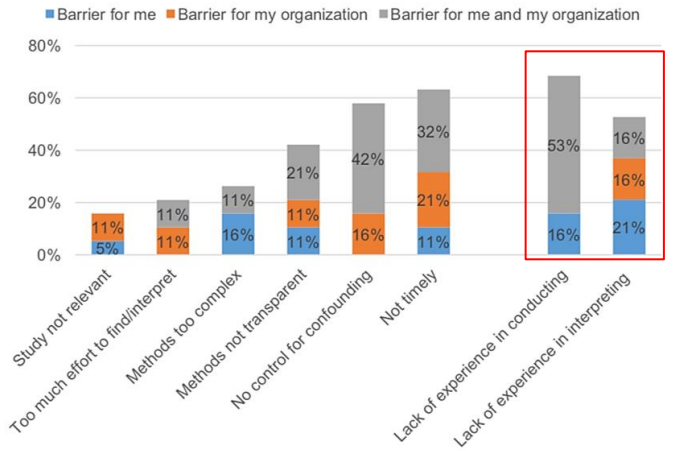
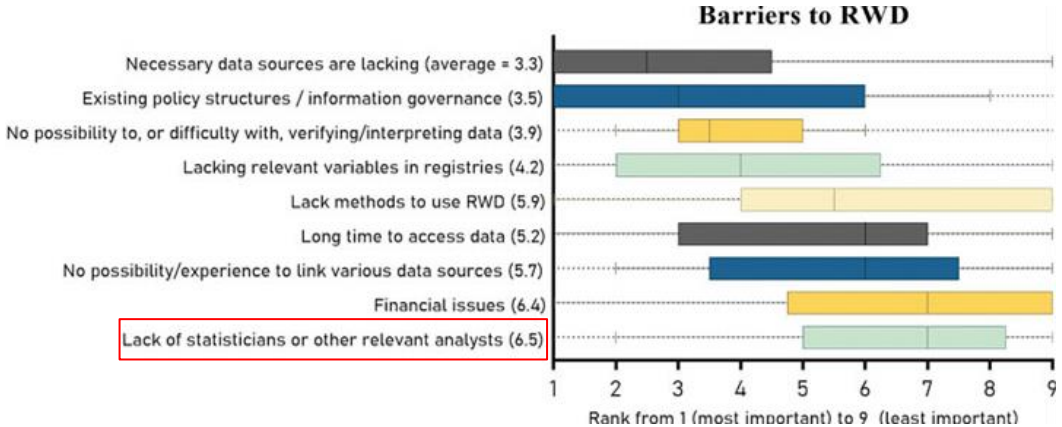


Fig. 4 – Perceived barriers to use of observational studies in decision making (N = 19).

Malone, et al. 2018. Study is of US payers

Lack of personnel



Hogervorst et al. 2022. Survey of 22 EUnetHTA member HTA organizations.

Existing RWE assessment tools fall short

BMJ Open

How well can we assess the validity of non-randomised studies of medications? A systematic review of assessment tools

Elvira D'Andrea [1], Lydia Vinals,[2] Elisabetta Patorno [1], Jessica M. Franklin,[1] Dimitri Bennett [3,4], Joan A. Largent,[5] Daniela C. Moga,[6] Hongbo Yuan,[7] Xuerong Wen,[8] Andrew R. Zullo,[9,10] Thomas P. A. Debray [11,12], Grammati Sarri [13]

Evaluated 44 assessment tools for non-randomized studies

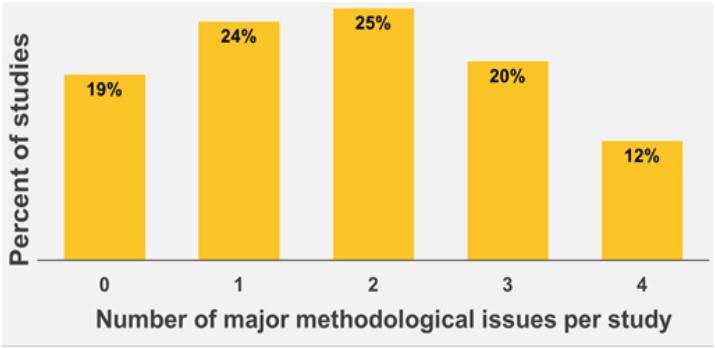
Conclusions:

- Most tools are primarily focused on reporting
- None covered all methodological domains

D'Andrea E, et al. How well can we assess the validity of non-randomised studies of medications? A systematic review of assessment tools. *BMJ Open*. 2021; 11(3):e043961

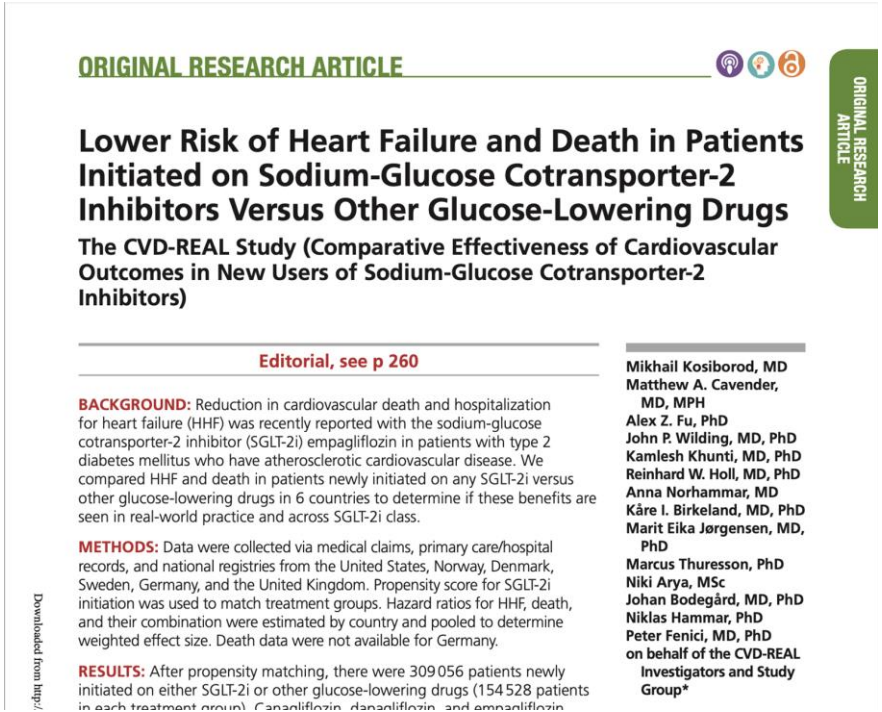
Prevalence of avoidable source of bias in published real-world studies of medication safety and effectiveness

Methodological characteristics	Total (N=75)	Cohort studies (N=65)	Case-control studies (N=10)
Non-user comparator – n (%)	41 (55)	31 (48)	10 (100)
Prevalent user design – n (%)	38 (51)	28 (43)	10 (100)
Time-related bias – n (%)	43 (57)	41 (63)	2 (20)
Adjustment for causal intermediaries without appropriate statistical models – n (%)	31 (41)	21 (32)	10 (100)
Depletion of susceptible individuals – n (%)	33 (44)	23 (35)	10 (100)
Reverse causation – n (%)	29 (39)	25 (38)	4 (40)
Residual confounding – n (%)	56 (75)	49 (75)	7 (70)
Lack of adjustment for confounders measured in claims – any of 1-6 below	54 (72)	47 (72)	7 (70)
1. Age	2 (3)	2 (3)	0 (0)
2. Gender	5 (7)	4 (6)	1 (10)
3. Indicating disease and severity	17 (23)	13 (20)	4 (40)
4. Comorbidities	2 (3)	2 (3)	0 (0)
5. Prior medication use	16 (21)	13 (20)	3 (30)
6. Prior healthcare utilization	26 (65)	44 (68)	5 (50)
Failure to evaluate or control confounders unavailable in claims	16 (21)	12 (18)	4 (40)
Detection bias – n (%)	16 (21)	11 (17)	5 (50)
Exposure misclassification – n (%)	23 (31)	21 (32)	2 (20)
Outcome misclassification – n (%)	8 (11)	7 (11)	1 (10)
Informative censoring – n (%)	6 (8)	6 (9)	0 (0)



Bykov K, et al. Prevalence of Avoidable and Bias-Inflicting Methodological Pitfalls in Real-World Studies of Medication Safety and Effectiveness. *Clin Pharmacol Ther*. 2022;111(1):209-217

Study that we will evaluate in this workshop



ORIGINAL RESEARCH ARTICLE

Lower Risk of Heart Failure and Death in Patients Initiated on Sodium-Glucose Cotransporter-2 Inhibitors Versus Other Glucose-Lowering Drugs

The CVD-REAL Study (Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors)

ORIGINAL RESEARCH ARTICLE

Editorial, see p 260

BACKGROUND: Reduction in cardiovascular death and hospitalization for heart failure (HHF) was recently reported with the sodium-glucose cotransporter-2 inhibitor (SGLT-2i) empagliflozin in patients with type 2 diabetes mellitus who have atherosclerotic cardiovascular disease. We compared HHF and death in patients newly initiated on any SGLT-2i versus other glucose-lowering drugs in 6 countries to determine if these benefits are seen in real-world practice and across SGLT-2i class.

METHODS: Data were collected via medical claims, primary care/hospital records, and national registries from the United States, Norway, Denmark, Sweden, Germany, and the United Kingdom. Propensity score for SGLT-2i initiation was used to match treatment groups. Hazard ratios for HHF, death, and their combination were estimated by country and pooled to determine weighted effect size. Death data were not available for Germany.

RESULTS: After propensity matching, there were 309 056 patients newly initiated on either SGLT-2i or other glucose-lowering drugs (154 528 patients in each treatment group). Canagliflozin, dapagliflozin, and empagliflozin

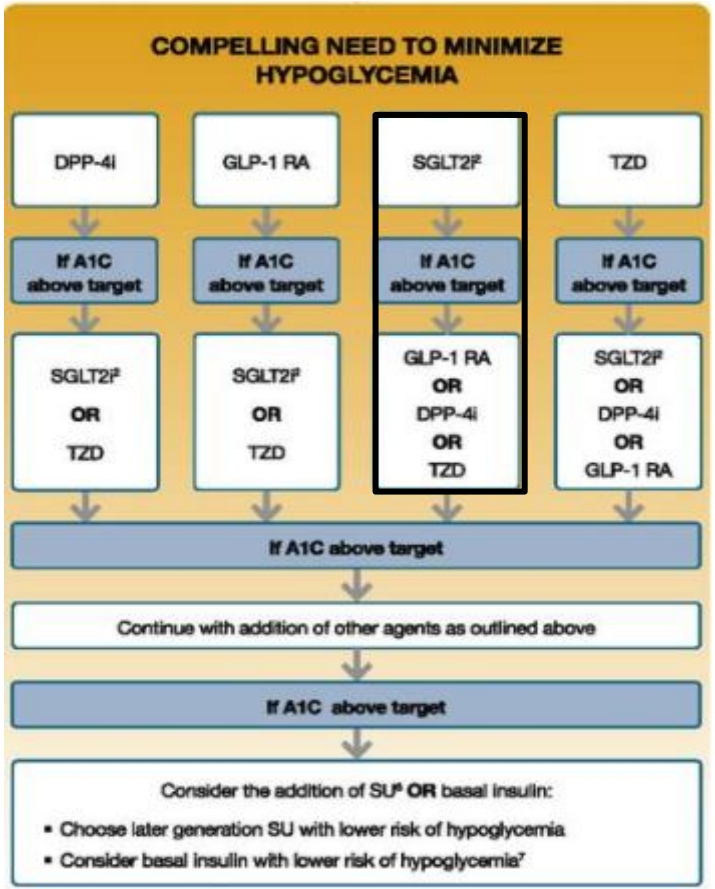
Mikhail Kosiborod, MD
Matthew A. Cavender, MD, MPH
Alex Z. Fu, PhD
John P. Wilding, MD, PhD
Kamlesh Khunti, MD, PhD
Reinhard W. Holl, MD, PhD
Anna Norhammar, MD
Kåre I. Birkeland, MD, PhD
Marit Eika Jørgensen, MD, PhD
Marcus Thuresson, PhD
Niki Arya, MSc
Johan Bodegård, MD, PhD
Niklas Hammar, PhD
Peter Fenici, MD, PhD
on behalf of the CVD-REAL Investigators and Study Group*

https://doi.org/10.1161/CIRCULATIONAHA.117.029190

Hypothetical HTA decision context

- The American Diabetes Association recommends multiple treatment options for 2L, including SGLT2s and GLP-1s.
- A manufacturer of an SGLT2 is seeking to become the preferred 2L treatment, especially in patients with cardiovascular risk.
- Assume the manufacturer submits this large RWE study of <300,000 patients as evidence of CV events as well as other evidence (e.g., EMPA-REG Outcomes RCT trial of one of the SGLT2s showed a substantial reduction in CV death and hospitalization for heart failure with one SGLT2).
- HTAs need to assess the validity of the RWE study to determine if this study can be used to support the manufacturer's clinical claims.

Note: Example uses a class of drugs, not a single treatment. Oversimplifies the evidence that would likely be submitted. Only meant to provide context for an HTA review..



American Diabetes Association. Pharmacological Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes - 2020. Excerpt from figure 9.1
https://care.diabetesjournals.org/content/43/Supplement_1/S98

Prerequisites before evaluating bias

- Pre-specified protocol with clearly defined comparative objectives, operationalized study elements and analysis plans
- Study elements† are relevant to HTA decision
- Sufficient data quality of study elements in RWD (fit for purpose) to allow interpretation of results
    - Adequate capture of each study element

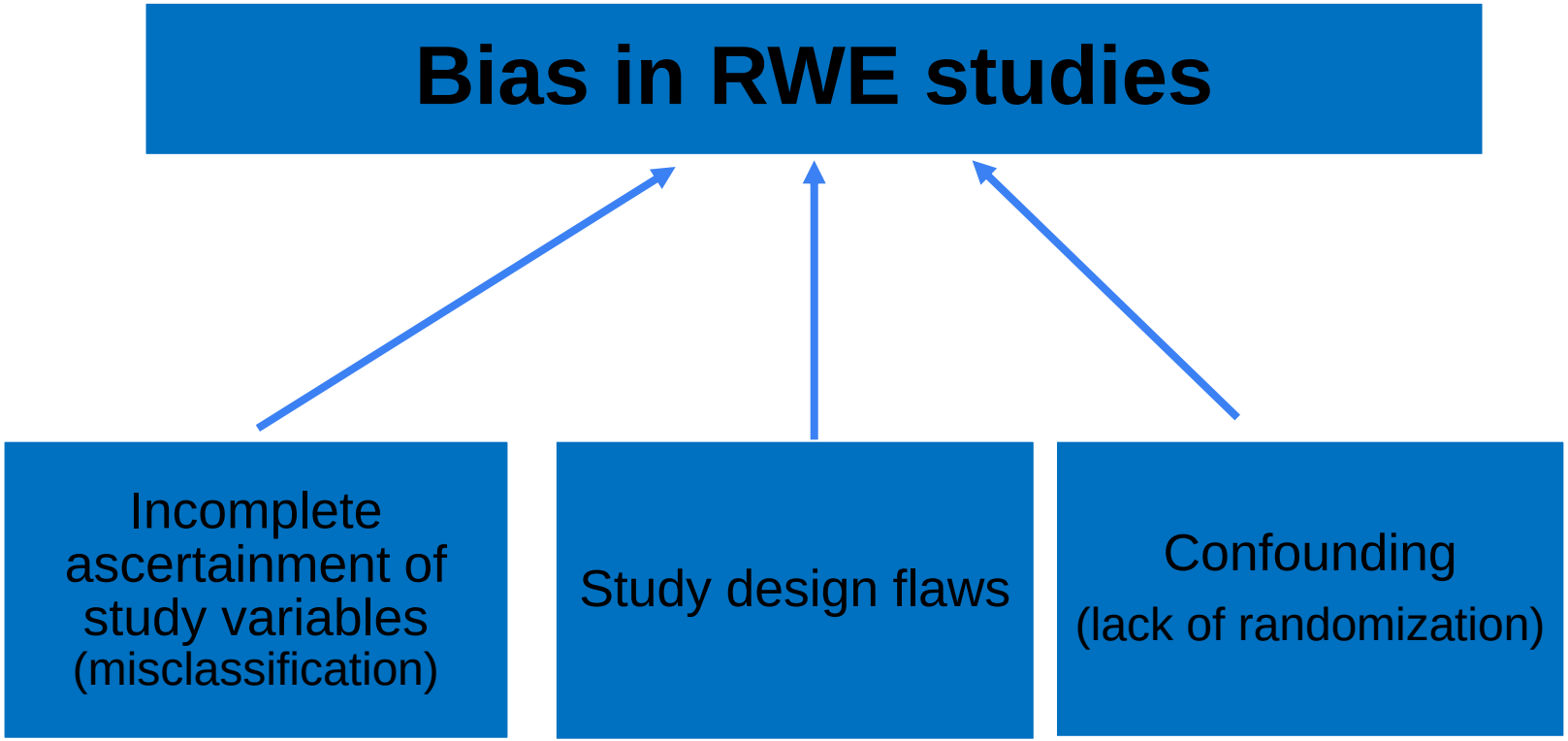
† Study elements: **P**opulation, **I**ntervention or treatment, **C**omparator, **O**utcome, study **T**ime or duration, **S**etting (PICOTS) plus key confounders

Study Evaluation

Katsiaryna Bykov, PharmD, ScD
Assistant Professor of Medicine

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine
Brigham and Women's Hospital, Harvard Medical School, Boston

Disclaimer: pharmacoepidemiologist conducting RWE studies in an academic center


Bias in RWE studies
Incomplete ascertainment of study variables (misclassification)
Study design flaws
Confounding (lack of randomization)

Tool domains

1. Study design
2-3. Exposure and outcome misclassification
4-9. Study design biases
10. Confounding
11. Propensity scores
12. Missing data
13. Summary

Study in a nutshell

Population:	T2DM
Intervention:	Sodium-glucose cotransporter-2 inhibitors (SGLT2i)
Comparator:	Other glucose-lowering drugs (oGLDs)
Outcomes:	Primary: hospitalizations for heart failure (HFF) Secondary: all-cause mortality; composite of mortality and HFF
Setting:	Deidentified health records from US, Norway, Denmark, Sweden, Germany, UK
Analysis:	SGLT2i initiators matched 1:1 on propensity score (PS) to oGLDs initiators; on-treatment (primary) and intention-to-treat (sensitivity) analyses were done

Exposure misclassification

- Is treatment/exposure assessment reliable?
  - Was medication use self-reported? Prescribing vs dispensing data?
  - Can it be obtained through other channels?

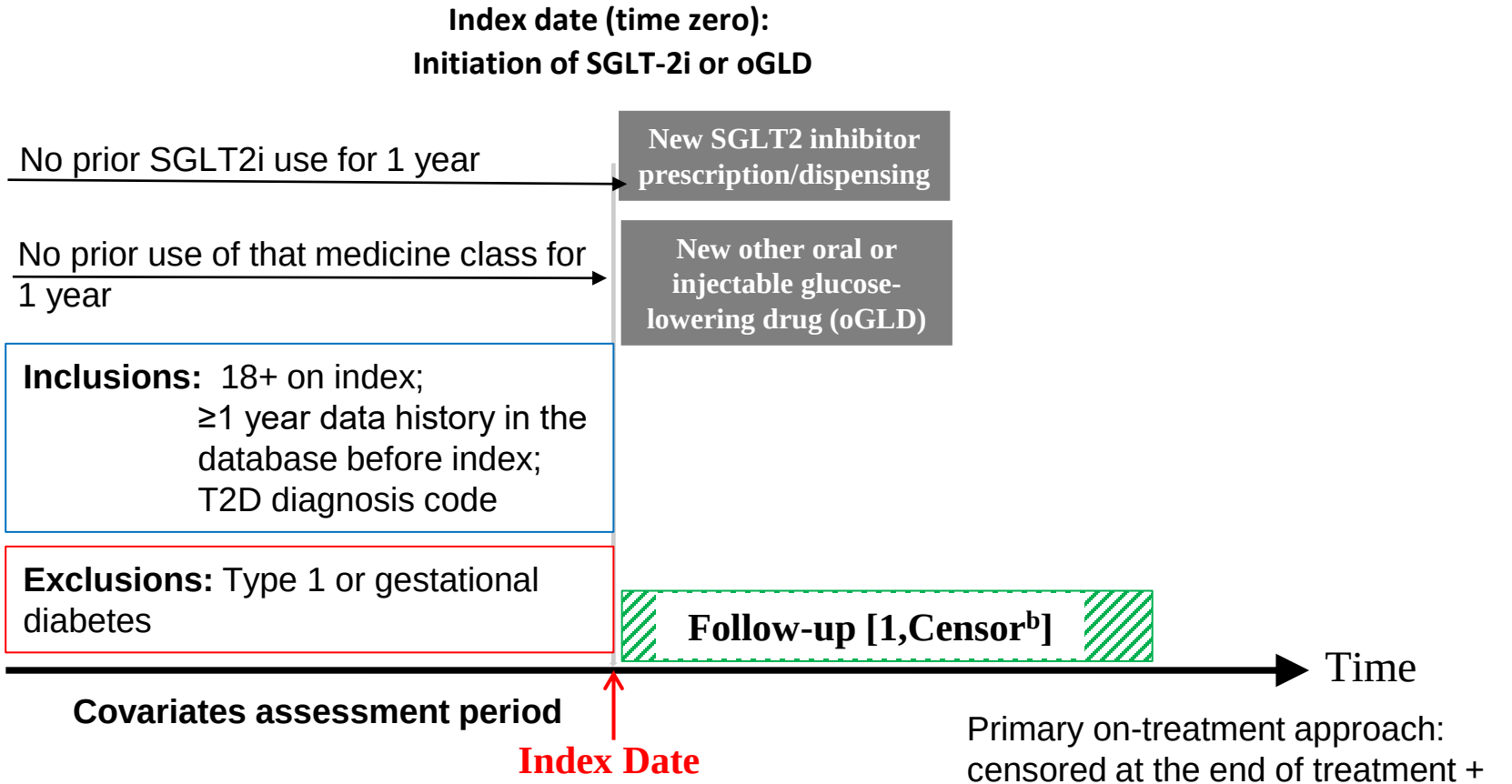
- Was only intention-to-treat analysis conducted but a high level of non-adherence expected?

Outcome misclassification

- Is outcome based entirely on disease codes without other kinds of information?
  - HHF: hospitalization required; primary discharge diagnosis
  - Mortality: based on Social Security Administration data linked to claims (US)
- If measures of performance (sensitivity, specificity, PPV, NPV) of outcome algorithms were provided, are there concerns they are not acceptable for the study population?
  - No
- Does the study evaluate incident (new) outcomes, but no outcome-free period before the start of follow-up was required?
  - No outcome-free period for HHF, but the outcome was hospitalization for HF (not incident HF)

Study design biases

- Time-related bias
- Inappropriate adjustment for causal intermediaries
- Depletion of outcome-susceptible individuals (selection bias)
- Reverse causation
- Detection bias
- Informative censoring



[b]**Censoring:** (i) End of index treatment; (ii) Outcome date; (iii) Migration/leaving practice/database; (iv) end of study period

Cohort study designs

Study design bias	Active comparator, new user	Active comparator, prevalent user	Non-user comparator, new user	Non-user comparator, prevalent user
Time-related bias			✓	✓
Depletion of outcome-susceptible individuals		✓		✓
Detection bias			✓	✓
Confounding			✓	✓

Time-related bias

- When eligibility for the study depends on measures collected after the beginning of follow-up (e.g., assessing type 2 diabetes during "any time during the study period" – involves 'looking into the future')
- When treatment assignment depends on exposure occurring after the beginning of follow-up (e.g., patients who ever used SGLT2i during study period start follow-up possibly before use; or requiring 2 prescriptions but one prescription can occur after start of follow-up)
- When individuals for any exposure group are selected first

Suissa S, Dell'Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2020

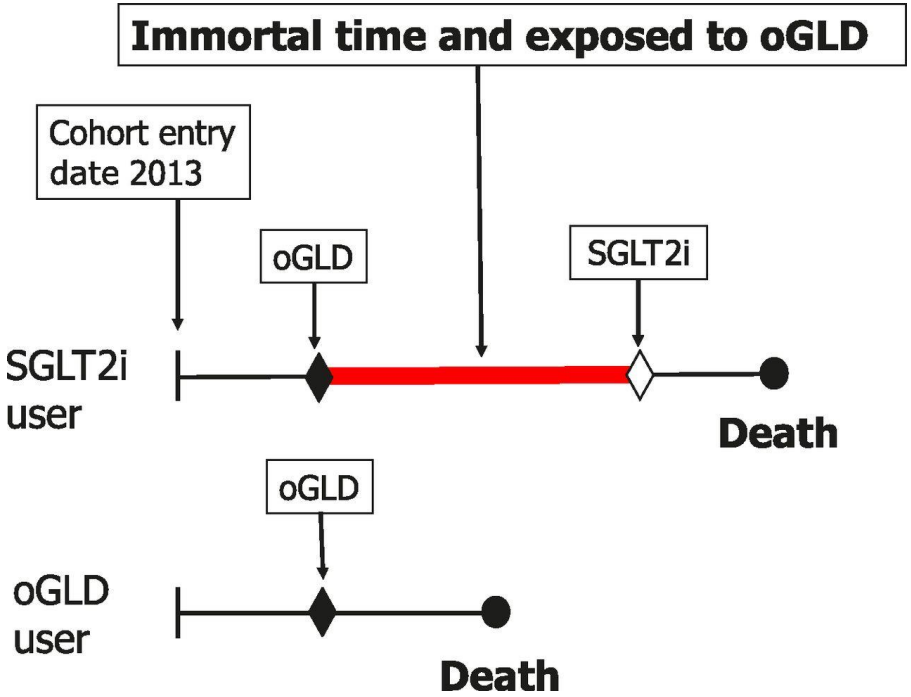
Time-related bias

Were inclusion or exclusion criteria measured during follow-up?	No
Was treatment group assignment based on exposure during follow-up?	No
Was SGLT2 inhibitor group identified first or selected preferentially?	Yes

Potential for time-related bias in this study?	**Yes**

From: **Lower Risk of Death With SGLT2 Inhibitors in Observational Studies: Real or Bias?**

Diabetes Care. 2017;41(1):6-10.



Depiction of immortal time bias: The top patient initiated treatment with an oGLD and subsequently switched to or added an SGLT2i, but the patient was classified as an SGLT2i user. The time between the first oGLD prescription and the first SGLT2i prescription is thus immortal (thick red line), since the subject must survive to receive this first SGLT2i prescription, but is not included as exposed to oGLD, leading to immortal time bias.

Inappropriate adjustment for causal intermediaries

- Were any covariates used for adjustment measured after the treatment decision was made/during follow-up
  AND no marginal structural models or G-methods were used?

  - all covariates were measured prior to treatment initiation

Depletion of outcome-susceptible individuals

- If comparison is to nonusers: did the study include prevalent (current) users, i.e. follow-up started NOT at drug initiation for all individuals?

- In an active-comparator, prevalent-user study, was start of follow-up aligned after treatment initiation?

Reverse causation

- Could outcomes have influenced the choice of treatment?

  - Were outcomes measured concurrent with treatment assessment?
  - Is recording of outcomes likely to be delayed?
  - Could early symptoms of undiagnosed outcomes have influenced treatment assignment?
  - Could treatment assignments have been influenced by procedures that are related to the outcomes?

Detection bias and informative censoring

- Detection bias: Are patients in one treatment group more likely to have the outcome recognized and recorded?

- Informative censoring: Were patients censored at treatment discontinuation without accompanying analyses to mitigate impact of informative censoring?

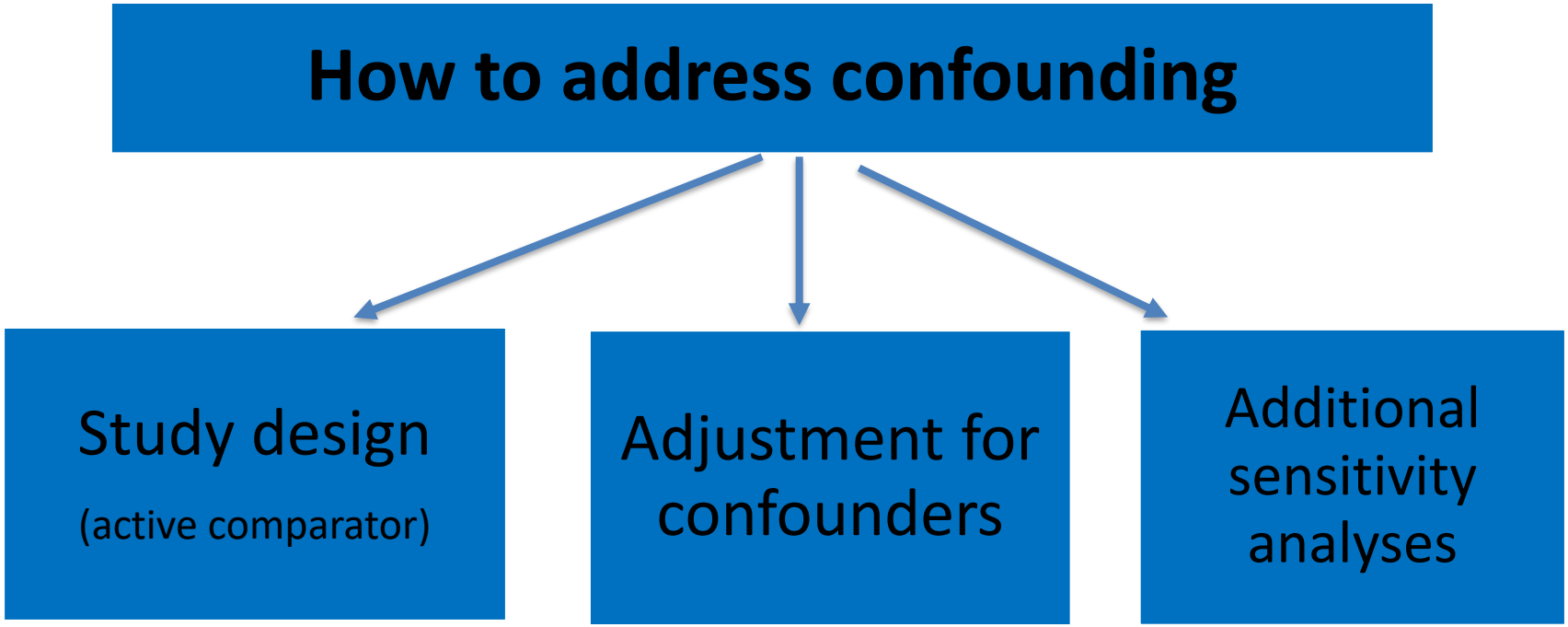
Detection bias and informative censoring

- Detection bias: Are patients in one treatment group more likely to have the outcome recognized and recorded?
  - Unlikely
- Informative censoring: Were patients censored at treatment discontinuation without analyses to mitigate impact of informative censoring?
  - ITT analysis was done
  - Grace period in on-treatment analysis

Study design biases

Potential for time-related bias	Yes
Inappropriate adjustment for causal intermediaries	No
Potential for depletion of outcome-susceptible individuals (selection bias)	No
Potential for reverse causation	No
Potential for detection bias	No
Potential for informative censoring	No

Confounding



Confounding

- Study design: active comparator ✓
  - **How similar is the active comparator to the treatment of interest?**

Confounding

- Study design: active comparator
  - Not sufficiently similar
- Adjustment for main confounders
  - Were main confounders included?
  - Yes
- Were additional analyses done to evaluate the impact of residual confounding?
  - No

POTENTIAL FOR IMPACTFUL RESIDUAL CONFOUNDING

Tool domains

1. Study design
2-3. Exposure and outcome misclassification
4-9. Study design biases
10. Confounding
11. Propensity scores
12. Missing data
13. Summary

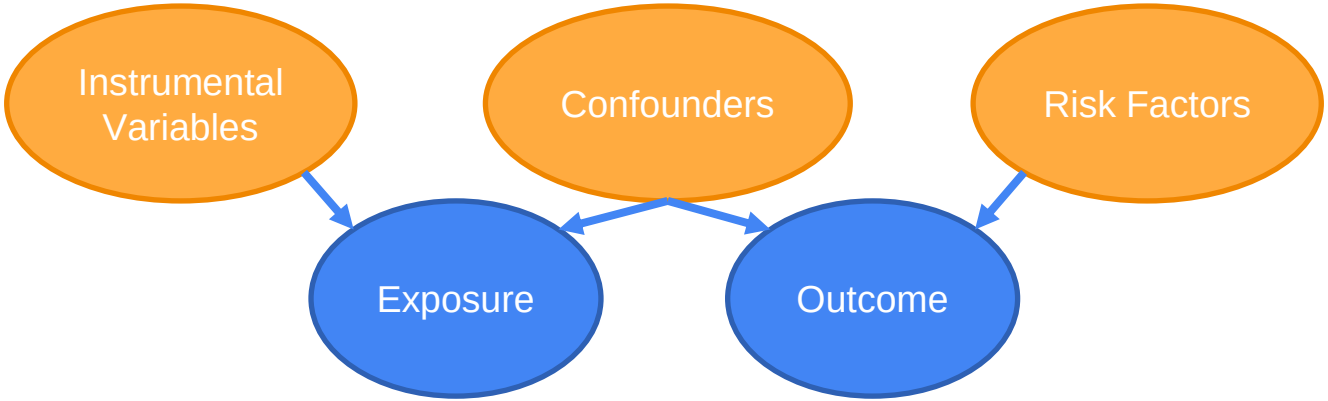
Optum

Statistical Methods Evaluation

Jessica M. Franklin, PhD
Principal Consultant, RWE
Optum Epidemiology

ISPOR, Boston, 2023

Propensity score model building



“A nonparsimonious propensity score was developed (separately within each country) for being initiated on an SGLT-2i to minimize confounding. Variables that may have affected treatment assignment or outcomes were included in the propensity score”

Variable selection methods

- Were all variables included in the model?
  - Yes
  - No use of stepwise or other approaches that select variables based on association with treatment
- Were other machine learning or automated approaches, such as the high-dimensional propensity score (hdPS), used for variable selection?
  - No
  - hdPS usually selects variables based on association with both outcome and exposure

Applying the PS

Based on propensity scores, patients receiving SGLT-2i were matched 1:1 with those receiving oGLDs. Nearest-neighbor caliper width of 0.25 multiplied by the SD of the propensity score distribution was used for the matching. In Sweden, Norway, and Denmark, an automated balance optimization method using the function Match (in package Matching) in R and a caliper of 0.2 were used for matching. The adequacy of propensity matching was assessed by standardized differences of postmatch patient characteristics. A significant imbalance was considered to be present if a >10% standardized difference was present between the 2 groups after propensity match.

PS diagnostics

- Was balance of covariates evaluated before applying the propensity score to the cohorts via weighting, matching, or some other approach?
  - Pre-matching balance available in the supplement
- Was balance of covariates evaluated after applying the propensity score to the cohorts via weighting, matching, or some other approach?

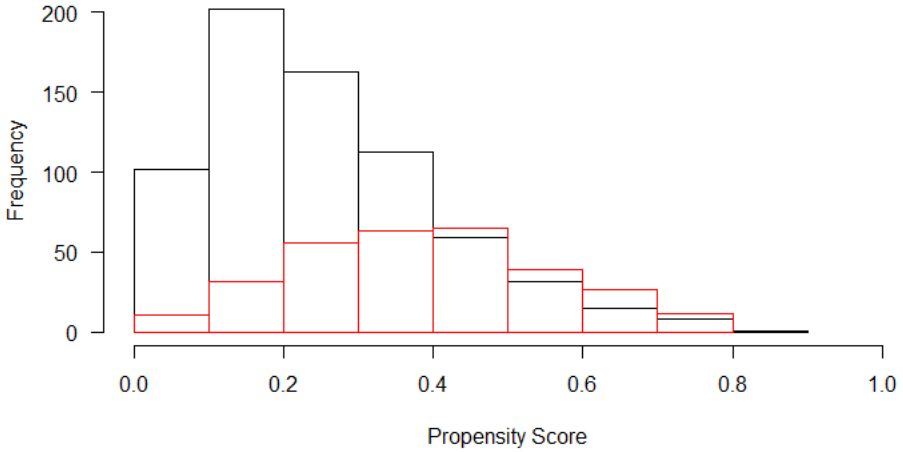
- Post-matching balance presented in the paper

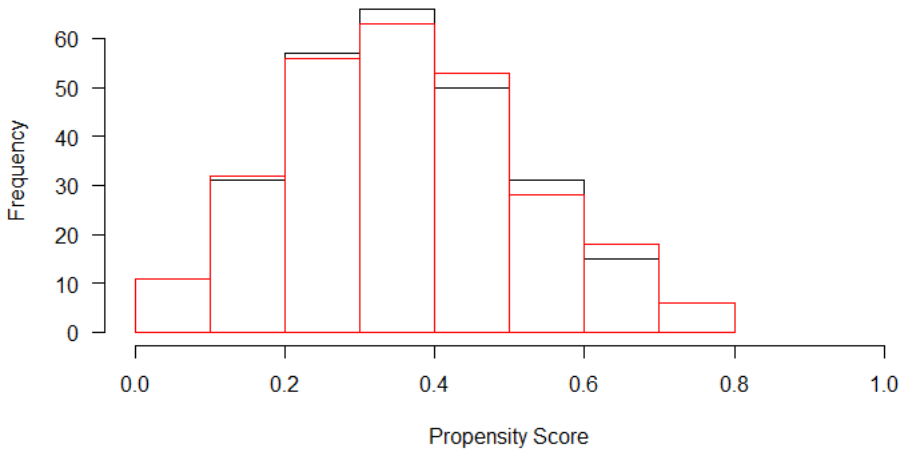
Table. Baseline Characteristics for All Countries Combined

	SGLT-2 Inhibitor (N=154 528)	Other GLD (N=154 528)
Mean age (SD), y	56.9 (10.0)	57.0 (10.6)
Women	68 420 (44.3)	68 772 (44.5)
Established cardiovascular disease*	20 044 (13.0)	20 302 (13.1)
Acute myocardial infarction	3793 (2.5)	3882 (2.5)
Unstable angina	2529 (1.6)	2568 (1.7)
Heart failure	4714 (3.1)	4759 (3.1)
Atrial fibrillation	5632 (3.6)	5698 (3.7)
Stroke	6337 (4.1)	6394 (4.1)
Peripheral arterial disease	5239 (3.4)	5229 (3.4)
Microvascular disease	42 217 (27.3)	42 215 (27.3)
Chronic kidney disease	3920 (2.5)	4171 (2.7)
Frailty (yes)†	11 982 (7.8)	12 731 (8.2)
Baseline glucose-lowering therapies		
Metformin	121 500 (78.6)	123 432 (79.9)
Sulfonylurea	59 406 (38.4)	59 788 (38.7)
Dipeptidyl peptidase-4 inhibitor	51 400 (33.3)	50 088 (32.4)
Thiazolidinedione	13 650 (8.8)	12 970 (8.4)
Glucagon-like peptide-1 receptor agonist	31 355 (20.3)	27 088 (17.5)
Insulin	45 573 (29.5)	45 097 (29.2)
Cardiovascular therapies		
Antihypertensive therapy‡	123 696 (80.0)	123 563 (80.0)
Loop diuretics	14 280 (9.2)	14 314 (9.3)
Thiazides	42 446 (27.5)	42 510 (27.5)
Angiotensin-converting enzyme inhibitors	66 812 (43.2)	67 067 (43.4)
Angiotensin receptor blockers	48 718 (31.5)	48 443 (31.4)
Statin therapy	103 968 (67.3)	104 128 (67.4)
Index year		
2012	21 (0.0)	270 (0.2)
2013	21 286 (13.8)	25 713 (16.6)
2014	71 070 (46.0)	58 793 (38.0)
2015	58 951 (38.1)	66 496 (43.0)

PS diagnostics

- Was the distribution of propensity scores evaluated separately by treatment group in the analytic samples (weighted or matched samples; potentially after trimming)?





Missing values in administrative data sources

- Primarily comprised of binary variables that describe whether specific diagnoses, procedures, or medications are observed or not observed.
- Study variables that are likely to have missing data include vital signs (blood pressure, BMI), laboratory measurements (HbA1c, LDL), and provider information (specialty).

Patient ID	Date	Diagnosis
12345	3/9/2023	Diabetes
12345	3/10/2023	Hypertension
12345	3/11/2023	Prior myocardial infarction

Patient ID	Date	Medication
12345	4/10/2023	5mg Amlodipine
12345	4/9/2023	500mg Metformin
12345	5/9/2023	10mg Dapagliflozin

Patient ID	Diabetes	Heart failure	CCBs	Statins
12345	1	0	1	0

Missing values in administrative data sources

- Primarily comprised of binary variables that describe whether specific diagnoses, procedures, or medications are observed or not observed.
- Study variables that are likely to have missing data include vital signs (blood pressure, BMI), laboratory measurements (HbA1c, LDL), and provider information (specialty).

Patient ID	Date	BMI
12345	11/10/2022	28.5
12345	1/10/2023	28.4
12345	4/10/2023	28.9

Patient ID	Baseline BMI (Prior 6 months)	Baseline BMI (7-12 months prior)
12345	28.6	--

Missing data

- Which study variables contain a significant proportion of missing values (> 5% of the population)?
  - Characteristics of US patients with and without vital status were similar (online-only Data Supplement Table III), indicating data missing completely at random because of administrative reasons.
- Was a complete case analysis conducted as the primary analysis?
  - Yes
- Does the likelihood of missing outcome information depend on patient characteristics that were not conditioned on in the analysis/outcome model?
  - No

Missing data in the outcome vs other variables

- If data are missing from the outcome only, a complete case analysis can be appropriate, as long as data are missing at random and all variables relevant to the missing data process are included in the outcome model.
- If data are missing from any other variables, complete case analysis is appropriate only if data are missing completely at random (missingness is not correlated with any study variables)

Missing data methods

- Other common approaches:
  - Missing indicator approach
  - Single imputation or last observation carried forward approach
  - Multiple imputation or inverse probability weighting to account for potential predictors of missingness
- If multiple imputation was conducted, were any variables that will be used in the analysis (outcome, exposure, and covariates) omitted in the imputation model?
- Does missingness depend on unmeasured factors (variables that are not captured in the data source)?

Summary

Cynthia J Girman, DrPH, FISPE
FOUNDER & PRESIDENT
www.CERobs.com

Disclosures:

CERobs Consulting - provides RWE & COA services to pharmaceutical companies
Member, PCORI Methodology Committee, ex-officio Clinical Trials Advisory Panel
Adjunct Professor, Dept of Epidemiology, University of North Carolina

SUMMARY: Potential for Bias

Pre-requisites: √pre-specified, √relevant to HTA decision, √assume sufficient data quality

Study design type	1a. Active comparator 1b. **Prevalent new user**	Higher risk of bias due to confounding-prevalent new user
Misclassification	2. Potential for exposure misclassification	
	3. Potential for outcome misclassification	
Study design biases	4. Potential for time-related bias **Selected SGLT2 first**	Major source of bias–often due to how patient are selected for study
	5. Inappropriate adjustment for causal intermediaries	
	6. Depletion of outcome-susceptibles/ selection bias	
	7. Bias due to Reverse causation	
	8. Detection bias	
	9. Bias due to Informative censoring	

SUMMARY: Potential for Bias (Continued)

Study Confounding	10. Potential for impactful residual confounding	All non-randomized studies have this potential; can explore the impact of residual confounding
	11. Potential for residual confounding due to suboptimal PS implementation or diagnostics	
Missing data	12. Potential for bias due to missing data	

CONCLUSION

- Questionnaire allows evaluating potential bias of a comparative study
  - Does not necessarily mean that the bias impacts interpretation
  - Must be considered in light of the totality of evidence about a product
  - Balance of practicality, scientific rigor, and potential impact of bias on results
- Difficult to determine which biases make a study uninterpretable
  - Direction of bias
  - Which are fatal flaws and which can be ignored
  - Varies from study to study
- Interpretation of RWE in presence of potential bias is **contextual**
  - Research objective
  - HTA decision
  - Magnitude of treatment effect
  - Results of pre-specified sensitivity or quantitative bias analysis

Q&A

For additional information or if you would like to be a pilot tester please reach out to:

Ashley Jaksa: ashley.jaksa@aetion.com
Katsiaryna (Kate) Bykov: kbykov@bwh.harvard.edu