Quantifying Potential Selection Bias in Observational Research: Simulations and Analyses Exploring **Religion and Depression** Using a Prospective UK Cohort Study (ALSPAC)

Jimmy Morgan

University of Bristol, UK

25/09/2024

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Selection Bias

- A major problem in research is non-random study participation, either from recruitment into the study or via loss-to-follow-up.
- This results in the analytic sample (i.e., those included in analyses) differing from the target population (i.e., the population of interest) (1).
- This can lead to selection bias and incorrect causal inferences (2).
- For instance, if participation is related to both the exposure and outcome, collider bias can occur.
- While selection bias is known to be an issue, the extent to which it can bias results is often difficult to know.

ALSPAC

- The Avon Longitudinal Study of Parents and Children is longitudinal cohort study based in Bristol, UK.
- ALSPAC recruited around 15,000 pregnant women and their offspring in the early 1990s and have followed up for the last 30 years.
- Selection bias has been explored in ALSPAC previously, with maternal age, ethnicity, sex, SEP and mental health all being associated with participation.

Rationale

- Our first paper found an association between attendance at a place of worship and continued participation in ALSPAC (3).
- While selection bias is often known to be an issue, the extent to which it can bias results is often difficult to know.
- We used the question “Does religiosity cause depression?” as the motivating example due to both the exposure and outcome being previously linked to participation in ALSPAC.

Methods

- We used a simulation study to explore how different patterns of selection bias (informed by realistic parameters) may impact results.

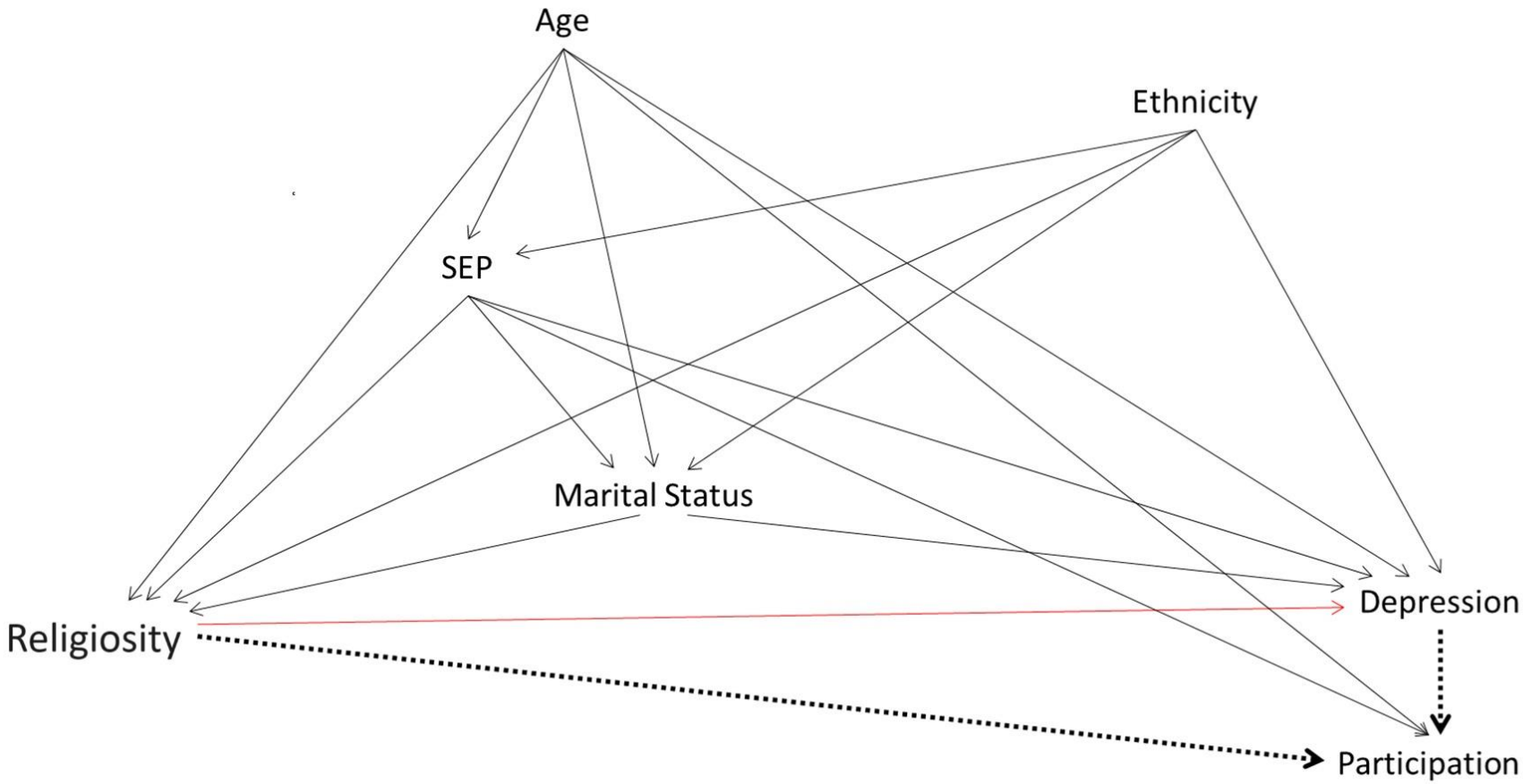
- We followed the ADEMP guidelines outlined by Morris et al., 2019 (4).
  - Aims
  - Data Generating Mechanism
  - Estimands
  - Methods
  - Performance Measures

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Methods - Continued

- **Aims:** To assess the bias caused by selection that may arise when estimating the association between religiosity and depression.
- **DGM:** Using parameters based on ALSPAC data, we explored causes and strength of selection, to investigate the magnitude and direction of bias.
- **Estimands:** When depression was binary, we used the log-odds estimate and when it was continuous we used the mean difference coefficient.
- **Methods:** 36 different selection scenarios varying the coding of the outcome, whether exposure caused outcome, exposure caused selection, and outcome caused selection.
- **Performance Measures:** We estimated both bias (how much the effect estimate differed from the true value) and coverage (proportion of simulations where the 95% CI included the true value) over 1,000 simulations.

DAG

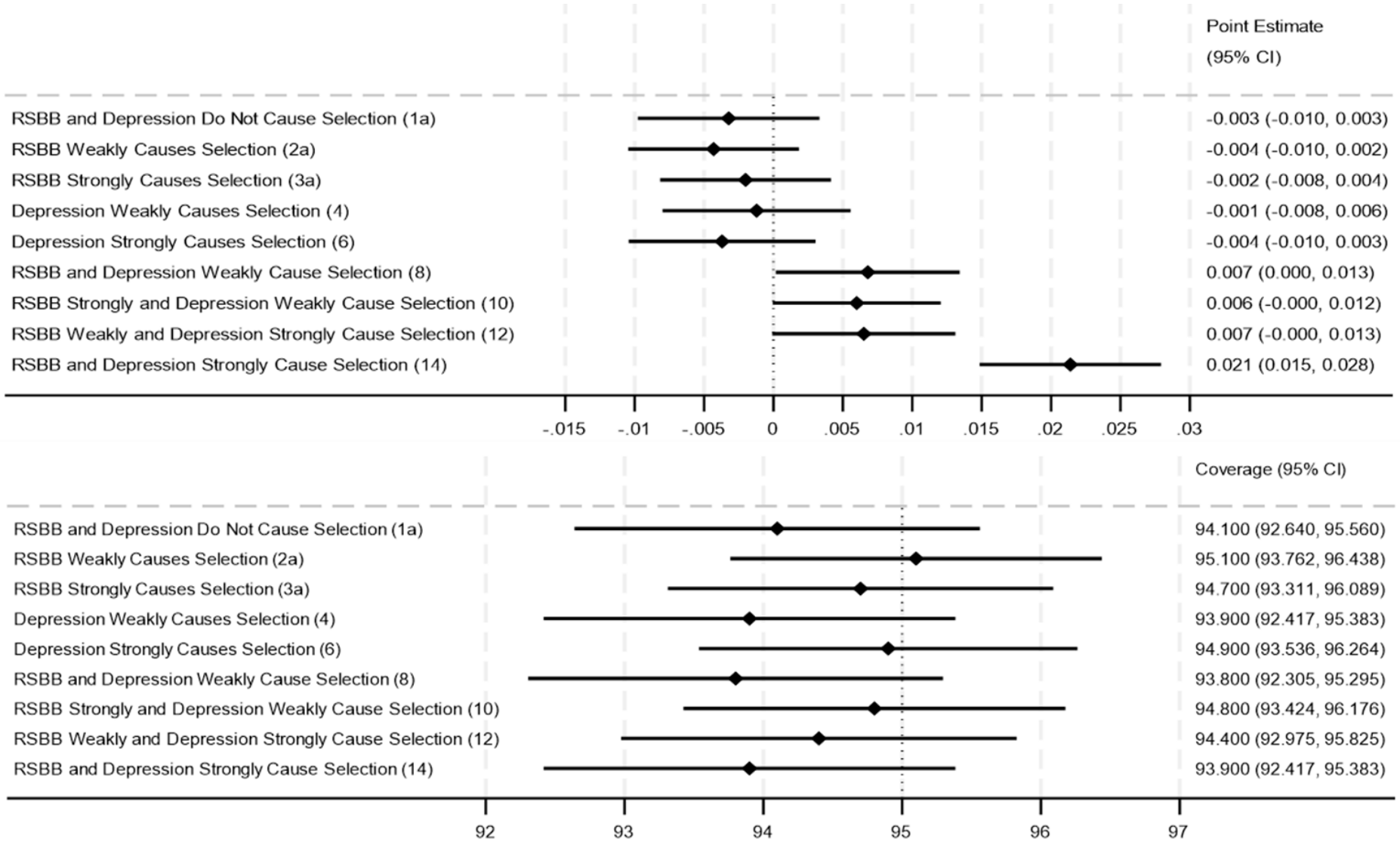


WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

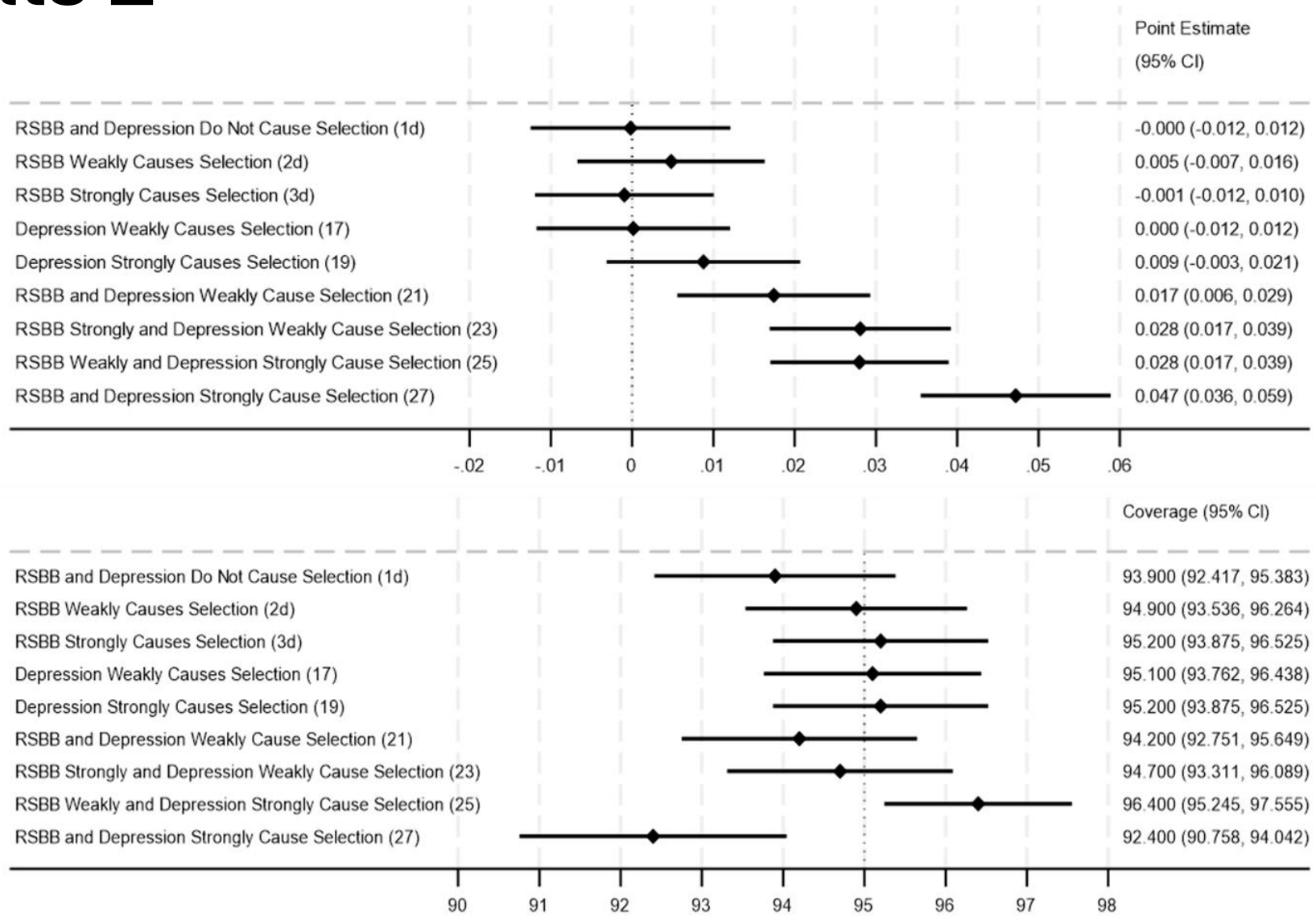
Missingness Models

Missingness Model	RSBB causes depression	Depression variable coding	Exposure (RSBB) causes selection	Outcome (depression) causes selection	Expected bias in adjusted analyses
1a	No	Binary	No	No	No
1b	No	Continuous	No	No	No
2b	No	Continuous	Yes (Weakly)	No	No
2c	Yes	Binary	Yes (Weakly)	No	No
3c	Yes	Binary	Yes (Strongly)	No	No
3d	Yes	Continuous	Yes (Strongly)	No	No
4	No	Binary	No	Yes (Weakly)	No
13	No	Continuous	Yes (Weakly)	Yes (Strongly)	Yes
16	Yes	Binary	No	Yes (Weakly)	No
22	Yes	Binary	Yes (Strongly)	Yes (Weakly)	Yes
27	Yes	Continuous	Yes (Strongly)	Yes (Strongly)	Yes

Results


Point Estimate
(95% CI)
RSBB and Depression Do Not Cause Selection (1a)
RSBB Weakly Causes Selection (2a)
RSBB Strongly Causes Selection (3a)
Depression Weakly Causes Selection (4)
Depression Strongly Causes Selection (6)
RSBB and Depression Weakly Cause Selection (8)
RSBB Strongly and Depression Weakly Cause Selection (10)
RSBB Weakly and Depression Strongly Cause Selection (12)
RSBB and Depression Strongly Cause Selection (14)
-0.003 (-0.010, 0.003)
-0.004 (-0.010, 0.002)
-0.002 (-0.008, 0.004)
-0.001 (-0.008, 0.006)
-0.004 (-0.010, 0.003)
0.007 (0.000, 0.013)
0.006 (-0.000, 0.012)
0.007 (-0.000, 0.013)
0.021 (0.015, 0.028)
-.015
-.01
-.005
0
.005
.01
.015
.02
.025
.03
Coverage (95% CI)
RSBB and Depression Do Not Cause Selection (1a)
RSBB Weakly Causes Selection (2a)
RSBB Strongly Causes Selection (3a)
Depression Weakly Causes Selection (4)
Depression Strongly Causes Selection (6)
RSBB and Depression Weakly Cause Selection (8)
RSBB Strongly and Depression Weakly Cause Selection (10)
RSBB Weakly and Depression Strongly Cause Selection (12)
RSBB and Depression Strongly Cause Selection (14)
94.100 (92.640, 95.560)
95.100 (93.762, 96.438)
94.700 (93.311, 96.089)
93.900 (92.417, 95.383)
94.900 (93.536, 96.264)
93.800 (92.305, 95.295)
94.800 (93.424, 96.176)
94.400 (92.975, 95.825)
93.900 (92.417, 95.383)
92
93
94
95
96
97

Results 2


Point Estimate
(95% CI)
RSBB and Depression Do Not Cause Selection (1d)
RSBB Weakly Causes Selection (2d)
RSBB Strongly Causes Selection (3d)
Depression Weakly Causes Selection (17)
Depression Strongly Causes Selection (19)
RSBB and Depression Weakly Cause Selection (21)
RSBB Strongly and Depression Weakly Cause Selection (23)
RSBB Weakly and Depression Strongly Cause Selection (25)
RSBB and Depression Strongly Cause Selection (27)
-0.000 (-0.012, 0.012)
0.005 (-0.007, 0.016)
-0.001 (-0.012, 0.010)
0.000 (-0.012, 0.012)
0.009 (-0.003, 0.021)
0.017 (0.006, 0.029)
0.028 (0.017, 0.039)
0.028 (0.017, 0.039)
0.047 (0.036, 0.059)
-.02
-.01
0
.01
.02
.03
.04
.05
.06
Coverage (95% CI)
RSBB and Depression Do Not Cause Selection (1d)
RSBB Weakly Causes Selection (2d)
RSBB Strongly Causes Selection (3d)
Depression Weakly Causes Selection (17)
Depression Strongly Causes Selection (19)
RSBB and Depression Weakly Cause Selection (21)
RSBB Strongly and Depression Weakly Cause Selection (23)
RSBB Weakly and Depression Strongly Cause Selection (25)
RSBB and Depression Strongly Cause Selection (27)
93.900 (92.417, 95.383)
94.900 (93.536, 96.264)
95.200 (93.875, 96.525)
95.100 (93.762, 96.438)
95.200 (93.875, 96.525)
94.200 (92.751, 95.649)
94.700 (93.311, 96.089)
96.400 (95.245, 97.555)
92.400 (90.758, 94.042)
90
91
92
93
94
95
96
97
98

Secondary Analysis

- We compared the religiosity-depression relationship at baseline and the most recent timepoints.
- Comparing these results should give us an idea of the maximum extent of selection bias in this relationship.
- Results largely followed the same pattern as the simulation study, with broadly similar results in both cohorts.

Strengths and Limitations

- Use of simulated data to create otherwise impossible scenarios.
- 36 different simulation scenarios.
- Models based on realistic values – Use of ALSPAC.

- ALSPAC is an almost perfect test case.
- Only looked at type 1 selection bias.
- Selection on enrolment into the study.

Summary

- Even when the exposure and outcome are related to participation, this does not always mean there will be substantial selection bias.
- Hopefully make future research more confident in conclusions drawn from ALSPAC religion data.
- Results apply to similar effect sizes and levels of missingness.

University of BRISTOL

CHILDREN OF THE 90s

Thank you for listening

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

JOHN TEMPLETON FOUNDATION

References

1. Lu H, Cole SR, Howe CJ, Westreich D. Toward a Clearer Definition of Selection Bias When Estimating Causal Effects. Epidemiology. 2022 Sep;33(5):699–706.
2. Hernan MA, Robins JM. Causal inference. Boca Raton, Fla.; London: CRC : Taylor & Francis [distributor; 2020.]
3. Morgan J, Halstead I, Northstone K, Major-Smith D. Religious/spiritual beliefs and behaviours and study participation in a prospective cohort study (ALSPAC) in Southwest England. Wellcome Open Res. 2022 Jul 12;7:186.
4. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019 May 20;38(11):2074–102.

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024