Causal inference in registry research

Magne Haugland Solheim
Øystein Ariansen Haaland
Copenhagen, January 2019

Health Registries for Research
Norway

NordForsk

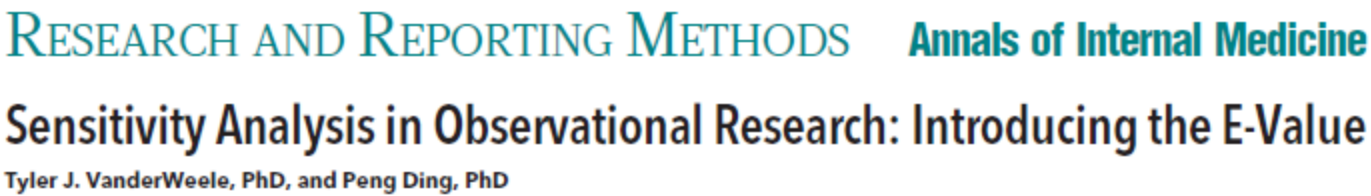
Outline

Sensitivity analysis

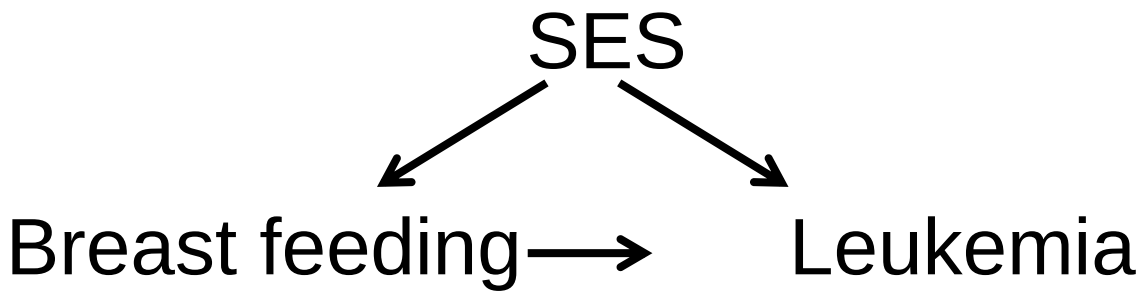
Bonus (if time)

Sensitivity analysis

Sensitivity analysis

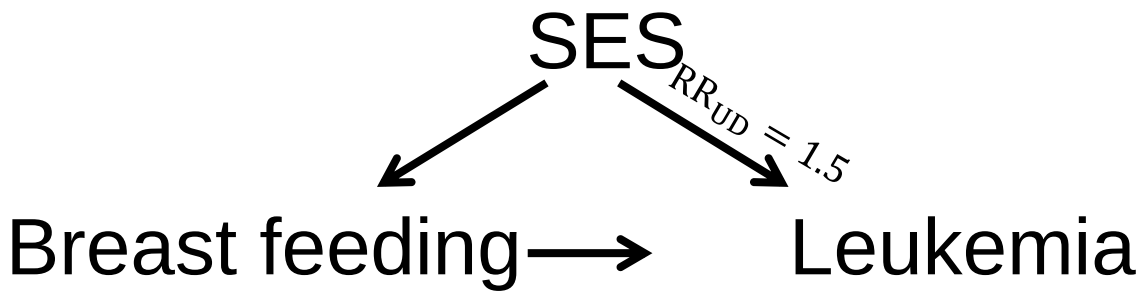


Sensitivity analysis



RR = 0.81 (0.71 – 0.91), p<0.001

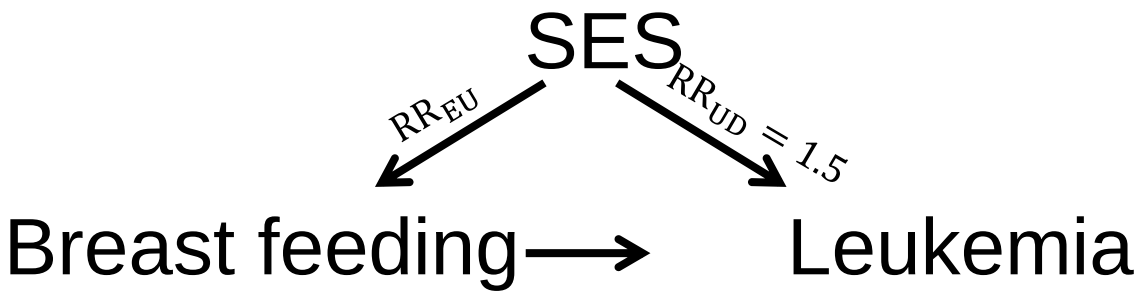
Sensitivity analysis



RR = 0.81 (0.71 – 0.91), p<0.001

RR_{UD}: Effect of SES on outcome

Sensitivity analysis

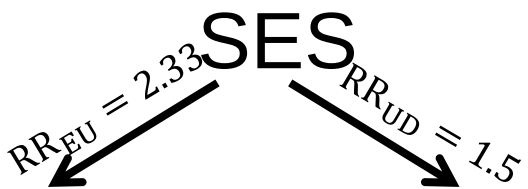


RR = 0.81 (0.71 – 0.91), p<0.001

RR_{UD}: Effect of SES on outcome

RR_{EU}: Imbalance in treatment group regarding SES

Sensitivity analysis



Breast feeding ⟶ Leukemia

RR = 0.81 (0.71 – 0.91), p<0.001

Low SES:

-Increases risk by 1.5 (RR_{UD})

-30% in the breast feeding group

-70% in the non-breast feeding group

-0.70/0.30=2.33 (RR_{EU})

Sensitivity analysis

Breast feeding ⟶ Leukemia

RR = 0.81 (0.71 – 0.91), p<0.001

Low SES:

-Increases risk by 1.5 (RR_{UD})

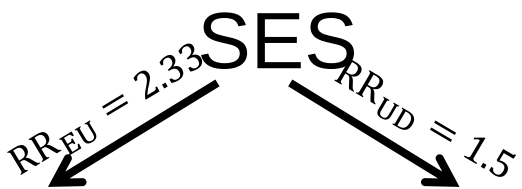
-30% in the breast feeding group

-70% in the non-breast feeding group

-0.70/0.30=2.33 (RR_{EU})

$$B = \frac{2.33 \times 1.5}{2.33 + 1.5 - 1} = 0.81$$

Sensitivity analysis



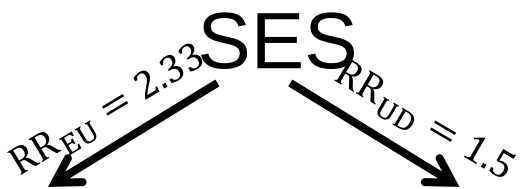
Breast feeding ⟶ Leukemia

RR = 0.81 (0.71 – 0.91), p<0.001

$$B = \frac{2.33 \times 1.5}{2.33 + 1.5 - 1} = 0.81$$

RR_{adj} = 0.81/B = 0.81/0.81 = 1.0

Sensitivity analysis



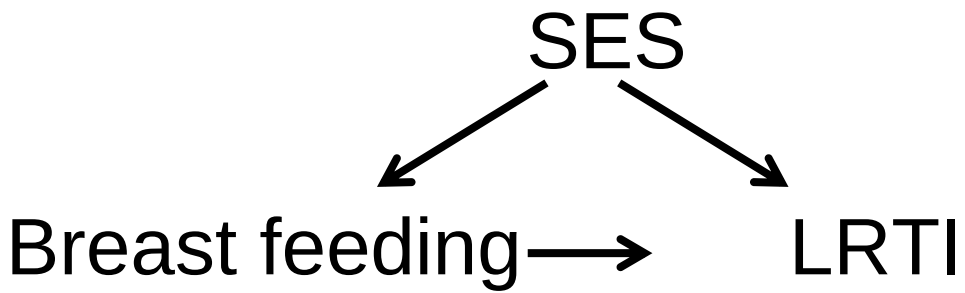
Breast feeding ⟶ Leukemia

RR = 0.81 (0.71 – 0.91), p<0.001

$$B = \frac{2.33 \times 1.5}{2.33 + 1.5 - 1} = 0.81$$

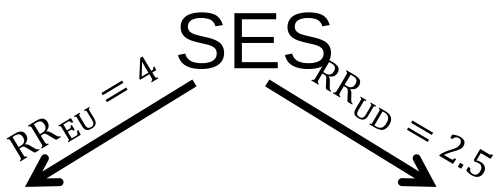
RR_{adj} (CI) = 1.0 (0.88 – 1.12)

Sensitivity analysis



RR = 0.28 (0.14 – 0.54), p<0.001

Sensitivity analysis



Breast feeding ⟶ LRTI

RR = 0.28 (0.14 – 0.54), p<0.001

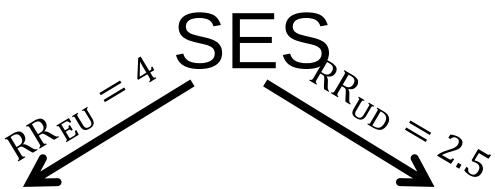
Low SES:

-Increases risk by 2.5

-20% in the breast feeding group

-80% in the non-breast feeding group

Sensitivity analysis



Breast feeding ⟶ LRTI

RR = 0.28 (0.14 – 0.54), p<0.001

Low SES:

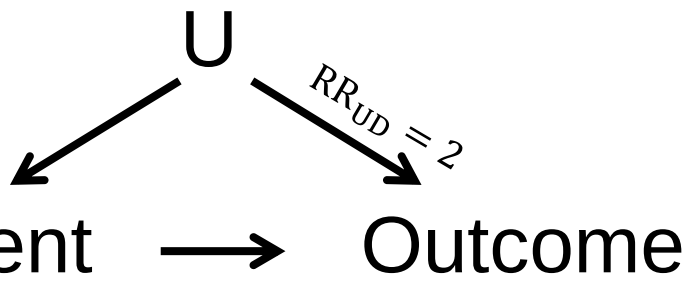
-Increases risk by 2.5

-20% in the breast feeding group

-80% in the non-breast feeding group

RR_{adj} (CI) = 0.51 (0.25 – 0.98)

Find B and adjust OR (CI) – I



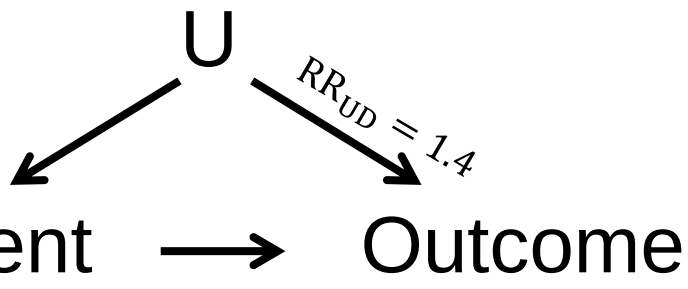
RR = 0.40 (0.30 – 0.51)

U:

-60% among unexposed

-80% among exposed

Find B and adjust OR (CI) – II



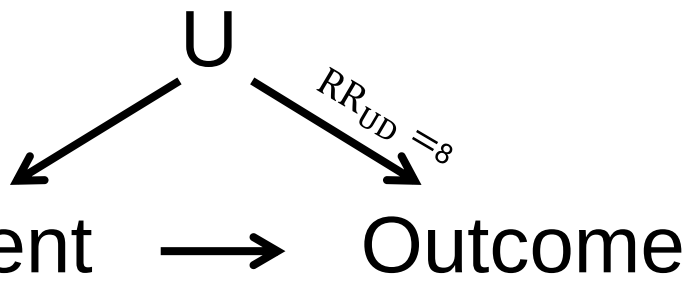
RR = 0.90 (0.89 – 0.91)

U:

-10% among unexposed

-20% among exposed

Find B and adjust OR (CI) – III



RR = 0.10 (0.04 – 0.17)

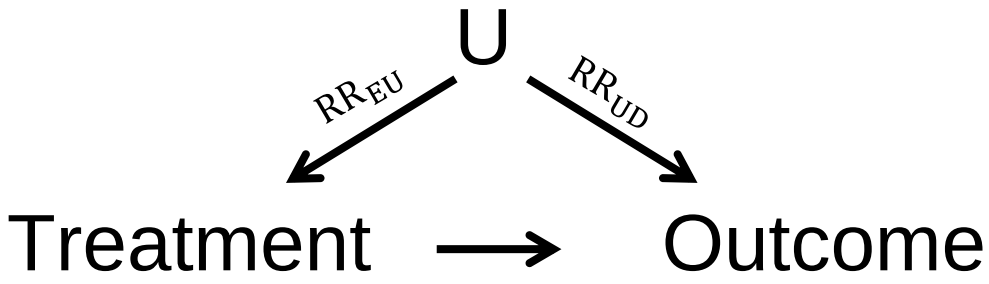
U:

-10% among unexposed

-50% among exposed

Sensitivity analysis

The E-value

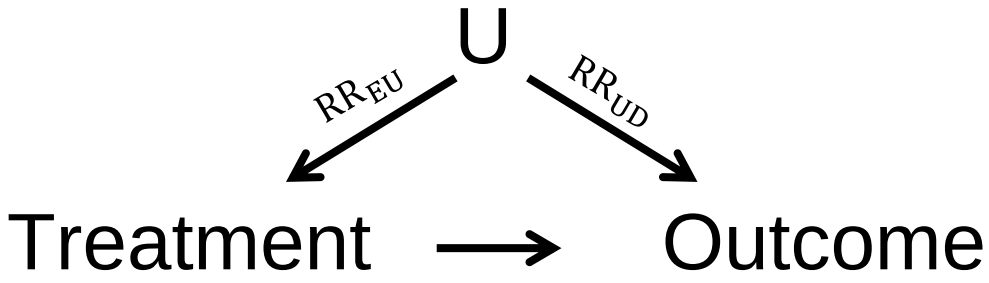


RR_{EU}: Imbalance in treatment group regarding U

RR_{UD}: Effect of U on outcome

Sensitivity analysis

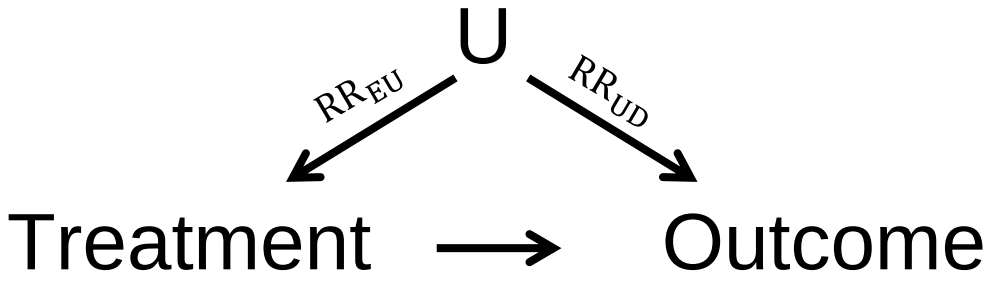
The E-value



E: Minimum value such that $RR_{EU} = RR_{UD}$, and $RR_{adj}=1 \Leftrightarrow B=RR$

Sensitivity analysis

The E-value

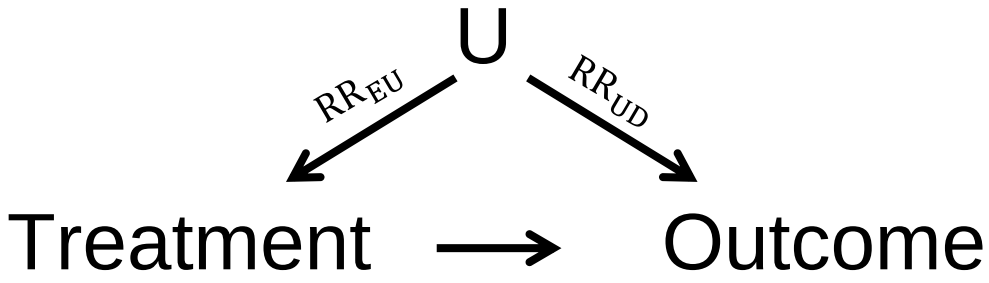


E: Minimum value such that $RR_{EU} = RR_{UD}$, and $RR_{adj}=1 \Leftrightarrow B=RR$

«The observed risk ratio of RR could be explained away by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of E each, above and beyond the measured confounders, but weaker confounding could not do so.»

Sensitivity analysis

The E-value

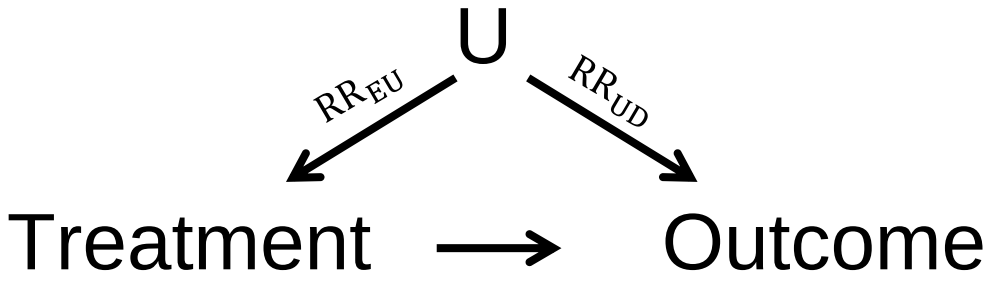


E: Minimum value such that $RR_{EU} = RR_{EU}$, and $RR_{adj}=1 \Leftrightarrow B=RR$

$$E = RR + \sqrt{RR \times (RR - 1)}$$

Sensitivity analysis

The E-value



E: Minimum value such that $RR_{EU} = RR_{EU}$, and $RR_{adj}=1 \Leftrightarrow B=RR$

$$E = RR + \sqrt{RR \times (RR - 1)}$$

Calculate both for RR and for part of CI that is closest to 1

Sensitivity analysis

Table 1. Calculating the E-Value for Risk Ratios

Estimate or CI, by Direction of Risk Ratio	Computation of the E-Value
***RR* >1**	
Estimate	E-value = $RR + \text{sqrt}\{RR \times (RR - 1)\}$
CI	If $LL \leq 1$, then E-value = 1 If $LL > 1$, then E-value = $LL + \text{sqrt}\{LL \times (LL - 1)\}$
***RR* <1**	
Estimate	Let $RR^* = 1/RR$ E-value = $RR^* + \text{sqrt}\{RR^* \times (RR^* - 1)\}$
CI	If $UL \geq 1$, then E-value = 1 If $UL < 1$, then let $UL^* = 1/UL$ and E-value = $UL^* + \text{sqrt}\{UL^* \times (UL^* - 1)\}$

LL = lower limit of the CI; *RR* = risk ratio; *RR** = inverse of *RR*; *UL* = upper limit of the CI; *UL** = inverse of *UL*.

Find E

RR = 0.40 (0.30 – 0.51)

RR = 0.90 (0.89 – 0.91)

RR = 0.10 (0.04 – 0.17)

RR = 10.0 (6 – 25)

RR = 3.1 (1.8 – 4.7)

Sensitivity analysis

Table 2. E-Values for Other Effect Measures

Effect Measure	Computation of Approximate E-Value
OR or HR for rare outcomes	When the outcome is relatively rare (e.g., <15%) by the end of follow-up, the E-value formula in Table 1 may be used (37). In a case–control study, the outcome only needs to be rare in the underlying population, not in the case–control study.
Rate ratio for count and continuous outcomes	For ratio measures for count outcomes (or nonnegative continuous outcomes), the E-value may be found by replacing the risk ratio with the rate ratio (or the ratio of expected values) in the E-value formula in Table 1 (37).
OR for common outcomes	When the outcome is common (>15% at the end of follow-up), an approximate E-value may be obtained by replacing the risk ratio with the square root of the OR (45), i.e., $RR \approx \text{sqrt(OR)}$, in the E-value formula in Table 1.
HR for common outcomes	When the outcome is common (>15% at the end of follow-up), an approximate E-value may be obtained (45) by applying the approximation $RR \approx (1 - 0.5^{\text{sqrt(HR)}})/(1 - 0.5^{\text{sqrt(1/HR)}})$ in the E-value formula in Table 1.
Difference in continuous outcomes	With standardized effect sizes d (mean of the outcome variable divided by the SD of the outcome) and an SE for this standardized effect size s_d, an approximate E-value may be obtained (45–47) by applying the approximation $RR \approx \exp(0.91 \times d)$ in the E-value formula. An approximate CI for the risk ratio may be found by using the approximation $(\exp\{0.91 \times d - 1.78 \times s_d\}, \exp\{0.91 \times d + 1.78 \times s_d\})$. This approach relies on additional assumptions and approximations. Other sensitivity analysis techniques have been developed for this setting (27–29), but they generally require additional assumptions, and the variables do not necessarily have a corresponding E-value.
Risk difference	If the adjusted risks for the treated and untreated are p_1 and p_0, then the E-value may be obtained by replacing the risk ratio with p_1/p_0 in the E-value formula. The E-value for the CI on a risk difference scale is more complex, and software to obtain this is described in the Supplement (available at Annals.org). Alternatively, if the outcome probabilities p_1 and p_0 are not very small or very large (e.g., if they are between 0.2 and 0.8), then the approximate approach for differences in continuous outcomes given previously may be used. Other sensitivity analysis techniques have been developed for this setting (27–29) but generally require additional assumptions and do not provide a corresponding E-value.

HR = hazard ratio; OR = odds ratio; *RR* = risk ratio.

Triangulation

NATURE | COLUMN: WORLD VIEW

Publish houses of brick, not mansions of straw

Papers need to include fewer claims and more proof to make the scientific literature more reliable, warns William G. Kaelin Jr.

23 May 2017

Triangulation

COMMENT · 23 JANUARY 2018

Robust research needs many lines of evidence

Replication is not enough. Marcus R. Munafò and George Davey Smith state the case for triangulation.

Marcus R. Munafò ✉ & George Davey Smith

Triangulation

TRIANGULATION — A CHECKLIST

- The different approaches address the same underlying question.
- The key sources of bias for each approach are explicitly acknowledged.
- For each approach, the expected directions of all key sources of potential bias are made explicit, where feasible.
- Ideally, some of the approaches being compared will have potential biases that are in opposite directions.
- Ideally, results from more than two approaches — which have different and unrelated key sources of potential biases — are compared. (Source: ref. 3)

Health Reg Norway

NordForsk

Bradford Hill criteria for causality

1. Strength (effect size). Stronger is better!
   E-value argues this as well.

2. Consistency (reproducibility). Consistent findings across methods and in different places.
   Triangulation approach.

3. Specificity. The more specific, the higher the likelihood of causality.
   Natural experiments.

4. Temporality. Effect after cause.

Bradford Hill criteria for causality

5. Biological gradient. Logical relationship between exposure and incidence.
   More exposure often yields higher incidence.

6. Plausibility. Can a plausible mechanism be proposed?
   Detailed biological mechanism are often not in the scope of register epidemiology.
   «Something genetic»

7. Coherence. Epidemiological and laboratory results should be similar.

8. Experiment. Experimental evidence can be useful when available.
   RCTs.

9. Analogy. «If this is bad, than that sould be bad as well.»

Thanks for listening!