STATISTICAL LABORATORY #1

What did the studies reveal?

Final discussion

Different groups, different random samples, one mechanism

Hosted by Nicola Orsini

November 6, 2026

The apparent contradiction

Group	RD_{overall}	$RD_{XY\|Z=0}$	$RD_{XY\|Z=1}$	Reversal?
1	+0.032	-0.095	-0.127	Yes
2	+0.051	-0.048	-0.115	Yes
3	+0.093	-0.029	-0.090	Yes
4	+0.033	-0.075	-0.206	Yes

What was common across the studies, and what differed?

Overall
Mortality appears higher in the drug group.

Within age groups
Mortality appears lower in the drug group.

One mechanism, different random samples

$$Z \sim \text{Bernoulli}(0.40)$$

$$X \mid Z = z \sim \text{Bernoulli}(0.30 + 0.40z)$$

$$Y \mid X = x, Z = z \sim \text{Bernoulli}(0.15 - 0.10x + 0.40z)$$

Variables

- $Z = 1$: older
- $X = 1$: drug
- $Y = 1$: death within five years

Conditional differences

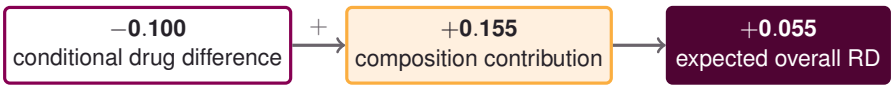
$$d_x = -0.10, \qquad d_z = +0.40$$

The mechanism is common; the observed counts vary from sample to sample.

The expected reversal

$$q_1 = P(Z = 1 \mid X = 1) = 0.609$$
$$q_0 = P(Z = 1 \mid X = 0) = 0.222$$
$$q_1 - q_0 = 0.387$$

$$\begin{aligned} RD_{\text{overall}} &= d_x + d_z(q_1 - q_0) \\ &= -0.100 + 0.40(0.387) \\ &= -0.100 + 0.155 \\ &= +0.055 \end{aligned}$$



When does the direction reverse?

$$C = d_z(q_1 - q_0)$$

1. Opposite directions

$$d_x C < 0$$

The conditional difference and the composition contribution point in opposite directions.

2. Composition is larger

$$|C| > |d_x|$$

The composition contribution is large enough to carry the overall difference through zero.

The reversal is surprising in the tables, but understandable from the mechanism.

Can the components predict the pilot results?

$$\widehat{d}_x = \widehat{RD}_{XY|Z=0}, \qquad \widehat{d}_z = \widehat{RD}_{ZY|X=0}$$

$$\widetilde{RD}_{\text{overall}} = \widehat{d}_x + \widehat{d}_z(\widehat{q}_1 - \widehat{q}_0)$$

Group	$\widehat{d}_x$	$\widehat{d}_z$	$\widehat{q}_1 - \widehat{q}_0$	Predicted RD_{overall}	Observed RD_{overall}	Gap
1	-0.095	+0.398	+0.369	+0.052	+0.032	−0.019
2	-0.048	+0.377	+0.368	+0.091	+0.051	−0.039
3	-0.029	+0.381	+0.415	+0.129	+0.093	−0.037
4	-0.075	+0.450	+0.410	+0.109	+0.033	−0.076

Gap = observed RD_{overall} minus predicted RD_{overall}.

The values should be close but need not be identical. In a finite sample, the two observed age-specific risk differences need not be equal.

What changes with a larger sample?

	Pilot, $n = 200$		Requested larger study		
Group	Observed	Predicted	n	Observed	Predicted
1	+0.032	+0.052	10,000	+0.072	+0.047
2	+0.051	+0.091	10,000	+0.055	+0.061
3	+0.093	+0.129	10,000	+0.046	+0.054
4	+0.033	+0.109	1,000	+0.024	+0.021

Values shown: RD_{overall}

Discuss

- Are estimates more similar across groups?
- Are they closer to $+0.055$?
- Is reversal more consistent?
- Are predicted and observed values closer?

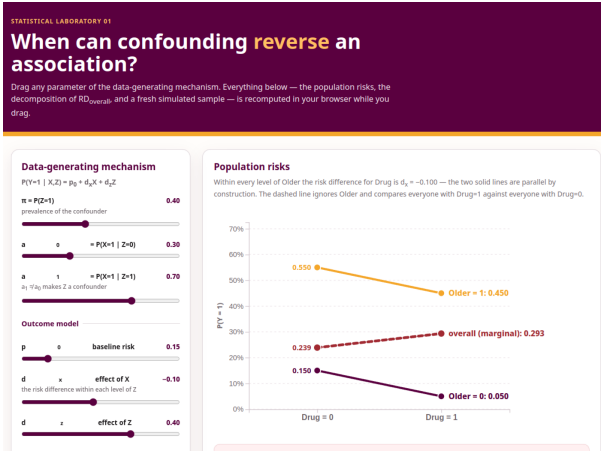
A larger sample does not change the expected mechanism. It reduces the variation of observed results around that expectation.

Take-home message

$$RD_{\text{overall}} = d_x + d_z(q_1 - q_0)$$

- Overall and conditional risk differences are different quantities.
- Reversal occurs when the composition contribution points in the opposite direction and outweighs the conditional difference.
- The mechanism determines what we expect.
- Sample size influences how much the observed results vary around that expectation.

Online simulator



stats4life.se/simulator/1/

Appendix: From overall to marginal RD

Under the additive mechanism,

$$p_{xz} = P(Y = 1 \mid X = x, Z = z) = p_0 + d_x x + d_z z.$$

Overall (crude) RD

Each drug group retains its own age distribution, $q_x = P(Z = 1 \mid X = x)$:

$$r_x = (1 - q_x)p_{x0} + q_x p_{x1}.$$

Therefore,

$$RD_{\text{overall}} = r_1 - r_0 = d_x + d_z(q_1 - q_0).$$

Standardized marginal RD

Use one common age distribution, with $q^* = P^*(Z = 1)$, under both drug conditions:

$$r_x^* = (1 - q^*)p_{x0} + q^* p_{x1}.$$

Therefore,

$$RD_{\text{marginal}} = r_1^* - r_0^* = d_x.$$

$q_1^* = q_0^* = q^*$, so $d_z(q_1^* - q_0^*) = 0$: common standardization removes the composition component.

Appendix: Three risk differences

Quantity	Comparison	Under the additive DGM
RD_{overall}	The observed drug groups, each retaining its own age distribution	$d_x + d_z(q_1 - q_0)$, the observed difference between groups.
$RD_{\text{standardized}}$	Both drug conditions averaged over the same age distribution	d_x: the composition contribution has been removed.
RD_{IPW}	Weighted pseudo-populations in which the drug groups have the same age distribution	The same marginal RD targeted by standardization.

Standardization and IPW make $q_1^* = q_0^*$, so $d_z(q_1^* - q_0^*) = 0$. Under this mechanism, $RD_{\text{standardized}} = RD_{\text{IPW}} = d_x$.

Precision. A larger study estimates the chosen target more precisely; it does not turn the crude contrast into the conditional treatment effect.

Effect modification. If the drug effect varies by age, the standardized and IPW RDs are weighted averages of the age-specific effects rather than one common d_x.