WEB MATERIAL

Relationship Between Level of American Football Playing and Diagnosis of Chronic Traumatic Encephalopathy in a Selection Bias Analysis

Jessica LeClair, Jennifer Weuve, Matthew P. Fox, Jesse Mez, Michael L. Alosco, Chris Nowinski, Ann McKee, and Yorghos Tripodis

Contents:

Web Appendix

Suppose we have the following observed data, where E and D represent exposure and disease levels, respectively:

	$E+$	$E-$
$D+$	a	b
$D-$	c	d

Furthermore, suppose we have the following data from our target population, which is unobserved:

	$E+$	$E-$
$D+$	A	B
$D-$	C	D

The true risk ratio of interest is:

$$RR = \frac{A/(A+C)}{B/(B+D)} = \frac{A(B+D)}{B(A+C)}.$$

We can express our observed counts as a weighted probability of being selected into the study, i.e.,

$$d = P(S|D-, E-) * D \rightarrow D = d * \frac{1}{P(S|D-, E-)}.$$

Substituting these selection probabilities into the RR equation:

$$RR_{Selection\ bias\ adjusted} = \frac{\frac{a/P(S|D+, E+)}{(a/P(S|D+, E+) + c/P(S|D-, E+))}}{\frac{b/P(S|D+, E-)}{(b/P(S|D+, E-) + d/P(S|D-, E-))}}$$

We want to factor out the RR for the observed data $\frac{a(b+d)}{b(a+c)}$:

$$= \frac{a/P(S|D+,E+)}{(a/P(S|D+,E+) + c/P(S|D-,E+))} * \frac{\frac{b}{P(S|D+,E-)} + \frac{d}{P(S|D-,E-)}}{b/P(S|D+,E-)}$$

$$= \frac{P(S|D+,E-)}{P(S|D+,E+)} \frac{a\left(\frac{b}{P(S|D+,E-)} + \frac{d}{P(S|D-,E-)}\right)}{b(\frac{a}{P(S|D+,E+)} + \frac{c}{P(S|D-,E+)})}$$

$$= \frac{P(S|D+,E-)}{P(S|D+,E+)} \frac{a\,(b+d)\frac{1}{(b+d)}\left(\frac{b\,P(S|D-,E-) + d\,P(S|D+,E-)}{P(S|D+,E-)P(S|D-,E-)}\right)}{b(a+c)\frac{1}{(a+c)}\left(\frac{a\,P(S|D-,E+) + c\,P(S|D+,E+)}{P(S|D+,E+)P(S|D-,E+)}\right)}$$

$$= \frac{P(S|D+,E-)P(S|D+,E+)P(S|D-,E+)}{P(S|D+,E+)P(S|D+,E-)P(S|D-,E-)} * \frac{a(b+d)}{b(a+c)} * \frac{(a+c)}{(b+d)}$$

$$* \frac{bP(S|D-,E-) + d\,P(S|D+,E-)}{a\,P(S|D-,E+) + c\,P(S|D+,E+)}$$

$RR_{Selection\ bias\ adjusted}$

$$= RR_{observed} \frac{(a+c)}{(b+d)} * \frac{P(S|D-,E+)}{P(S|D-,E-)} * \frac{b\,P(S|D-,E-) + d\,P(S|D+,E-)}{a\,P(S|D-,E+) + c\,P(S|D+,E+)}$$