β-amyloid example

```
# ##### Pike 2007
#
# # estimate SD of X from scraped data
# setwd("~/Dropbox/Personal computer/Independent studies/E-value analog (EVA)/Possible app
# d = read.csv( "scraped_pike.csv", header = FALSE )
#
# d = read.csv( "full_prepped_dataset.csv", header = FALSE )
#
# d = read.csv( "prepped_data.csv" )
#
# # use Table 1 to get marginal expectation of episodic memory
# N = c(31, 33, 32, 24, 6, 3, 20, 13)
# means = c(-3.29, -2.04, -0.02, -2.7, -0.1, -0.6, -2.7, -1.1)
# EY = sum( N * means ) / sum(N)
# bx = -0.67
#
# # assuming bivariate normality and no interactions
# ( E.pt = uniroot( f = function(x) ( EY + delta * bx - EY * x ) * ( (1+x) / (2*x) ), inte
# # confounding above and beyond age, sex, education
#
# # for confidence interval
# # they just report p < 0.001, so assume p = 0.001 exactly
# p = 0.001
# z = abs( qnorm(p/2) )
# SE = bx / z # because bx / SE = z
# bx.lo = bx - SE * abs( qnorm(0.025) )
#
# # can also scrape data from Fig 3 to check bivariate normality
#
#
# # plot the ACE bound
# ACE_bound = function( .EY, .delta, .bx, .B ) {
#   ( .EY + .delta * .bx - .EY * .B ) * ( (1+.B) / (2*.B) )
# }
#
# B = seq(1, 5, 0.5)
# ACE = ACE_bound( EY, delta, bx, B )
#
# plot( B, ACE )
#
# B = 1.1
# ( EY + delta * bx - EY * B )
# ( (1+B) / (2*B) )
```

Appendix - NEW VERSION

Converting univariable regression results to Cohen's d

Assume that (X, Y) are bivariate normal and consider the Cohen's d associated with an increase of Δ units in X:

$$
\begin{aligned}
d &= \frac{E[Y \mid X = c + \Delta] - E[Y \mid X = c]}{\sigma_{Y|X}} \\
&= \frac{\Delta \beta}{\sigma_Y \sqrt{1 - \rho_{XY}^2}} \\
&= \frac{\Delta \beta}{\sigma_Y \sqrt{1 - \frac{\beta^2 \sigma_X^2}{\sigma_Y^2}}} \\
&= \frac{\Delta \rho_{XY}}{\sigma_X \sqrt{1 - \rho_{XY}^2}}
\end{aligned}
$$

where the denominator comes from a well-known property of the bivariate normal distribution.

An approximate standard error can be derived using the delta method. Let $z_{fis} = XXX$ be the Fisher-transformed correlation, which is approximately normal with variance $\frac{1}{N-3}$. Define the transformation:

$$
\begin{aligned}
g(z_{fis}) = d &= \frac{\Delta \tanh(z_{fis})}{\sigma_X \sqrt{1 - \tanh^2(z_{fis})}} \\
SE_d &\approx \sqrt{\mathrm{Var}(z_{fis}) \left(g'(z_{fis}) \right)} \\
&= \frac{1}{\sqrt{N-3}} \times \frac{\Delta}{\sigma_X \sqrt{\mathrm{sech}^2(z_{fis})}} \\
&= \frac{\Delta}{\sigma_X \sqrt{(N-3)(1 - \rho_{XY}^2)}} \\
&= \frac{\Delta}{\sigma_X \sqrt{(N-3)\left(1 - \beta^2 \frac{\sigma_X^2}{\sigma_Y^2}\right)}}
\end{aligned}
$$

E-value for a univariable regression

$$
RR \approx \exp\left(0.91 \times \frac{\Delta \rho_{XY}}{\sigma_X \sqrt{1 - \rho_{XY}^2}} \right)
$$

CI:

$$RR_{lb} \approx \exp\left(0.91 \times \frac{\Delta\rho_{XY}}{\sigma_X\sqrt{1-\rho_{XY}^2}} - 1.78 \times SE_d\right)$$

$$= RR_{ub} \approx \exp\left(0.91 \times \frac{\Delta\rho_{XY}}{\sigma_X\sqrt{1-\rho_{XY}^2}} + 1.78 \times SE_d\right)$$

Converting multivariable regression results to Cohen's d

$$R_{Y\sim X|Z}^2 = 1 - \frac{SSE_{full}}{SSE_{red}}$$

$$= 1 - \frac{\sigma_{Y|X,\mathbf{Z}}^2}{\sigma_{Y|\mathbf{Z}}^2}$$

$$\sigma_{Y|X,\mathbf{Z}}^2 = \sigma_{Y|\mathbf{Z}}^2\left(1 - R_{Y\sim X|Z}^2\right)$$

DOES THIS USE MVN ASSUMPTION?

$$d = \frac{E[Y \mid X = c + \Delta, \mathbf{Z}] - E[Y \mid X = c, \mathbf{Z}]}{\sigma_{Y|X,\mathbf{Z}}}$$

$$= \frac{\Delta\beta}{\sigma_{Y|\mathbf{Z}}\sqrt{1 - R_{Y\sim X|Z}^2}}$$

$$\geq \frac{\Delta\beta}{\sigma_Y\sqrt{1 - R_{Y\sim X|Z}^2}}$$

Unlike in the univariable case, a simple relationship between β and $R_{Y\sim X|Z}^2$ is not available with additional distributional assumptions, so both quantities are needed to approximate d.

Standard error: Going to be hard because inference for $R_{Y\sim X|Z}^2$ won't be available

- describe what to do for preventive: just set delta to be negative?
- say that, for univariable case, if SD of X isn't available, can set Delta = SD of X

OLD STUFF

- outcome should be nonnegative: problem for bivariate normality...

Suppose we estimate the conditional mean of a nonnegative outcome Y given a continuous exposure of interest, X, and arbitrary measured covariates $\mathbf{Z}$ (taken to include any interactions modeled between X and other measured covariates), via a linear regression model:

$$\widehat{E}[Y \mid X, \mathbf{Z}] = \widehat{\beta}_0 + \widehat{\beta}_X X + \widehat{\boldsymbol{\beta}}_{\mathbf{Z}}' \mathbf{Z}$$

where $\widehat{\beta}_0$ denotes the estimated intercept, $\widehat{\beta}_X$ denotes the estimated coefficient of X, and $\widehat{\beta}_Z$ denotes a vector of estimated coefficients for $\mathbf{Z}$. To define an effect size on which to perform sensitivity analysis, fix a contrast

of interest between two levels of X (denoted $X = c_2$ vs. $X = c_1$). We take the causal effect of ultimate interest to be the average causal effect (ACE) in a hypothetical population in which the covariate distributions of those with $X = c_2$ in the sample and of those with $X = c_1$ in the sample each have prevalence 50%. Then, letting B denote the joint bias factor (VanderWeele and Ding 2017), we can perform sensitivity analysis as a special case of Ding and VanderWeele (2016)'s Proposition A.20:

$$
\begin{aligned}
ACE &\geq \left(\widehat{E}\big[Y \mid X = c_2\big] - \widehat{E}\big[Y \mid X = c_1\big] \cdot B \right) \times \left(\frac{1}{2} + \frac{1}{2B} \right) \\
&= \left(\widehat{E}_{\mathbf{Z}}\Big[\widehat{E}_Y\big[Y \mid X = c_2, \mathbf{Z}\big]\Big] - \widehat{E}_{\mathbf{Z}}\Big[\widehat{E}_Y\big[Y \mid X = c_1, \mathbf{Z}\big]\Big] \cdot B \right) \times \left(\frac{1}{2} + \frac{1}{2B} \right) \\
&= \left(\widehat{E}_{\mathbf{Z}}\big[\widehat{\beta}_0 + \widehat{\beta}_X c_2 + \widehat{\boldsymbol{\beta}}'_{\mathbf{Z}}\mathbf{Z} \big] - \widehat{E}_{\mathbf{Z}}\big[\widehat{\beta}_0 + \widehat{\beta}_X c_1 + \widehat{\boldsymbol{\beta}}'_{\mathbf{Z}}\mathbf{Z} \big] \cdot B \right) \times \left(\frac{B+1}{2B} \right) \\
&= \left(\big(\widehat{\beta}_0 + \widehat{\beta}_X c_2 + \widehat{\boldsymbol{\beta}}'_{\mathbf{Z}}\widehat{E}[\mathbf{Z}] \big) - \big(\widehat{\beta}_0 + \widehat{\beta}_X c_1 + \widehat{\boldsymbol{\beta}}'_{\mathbf{Z}}\widehat{E}[\mathbf{Z}] \big) \cdot B \right) \times \left(\frac{B+1}{2B} \right) \\
&= \left((1 - B)\big(\widehat{\beta}_0 + \widehat{\boldsymbol{\beta}}'_{\mathbf{Z}}\widehat{E}[\mathbf{Z}] \big) - \widehat{\beta}_X \left(c_2 - Bc_1 \right) \right) \times \left(\frac{B+1}{2B} \right)
\end{aligned}
\tag{1}
$$

The E-value for the point estimate can then be approximated by setting the left-hand side equal to 0 and using a numerical search to solve for B, and the E-values for the confidence interval can be approximated by replacing $\widehat{\beta}_X$ in the above expression with its lower confidence interval limit. The interpretations of B and the E-value change slightly given the continuous outcome (Ding and VanderWeele 2016): the sensitivity parameter RR_{UY} becomes a maximum ratio of conditional outcome means, rather than a maximum relative risk, while RR_{XU} is still interpreted in terms of the relative risk comparing levels of U conditional on $X = c_2$ versus $X = c_1$. Thus, an E-value of 2 would imply that true ACE defined above could be zero if there were, for example, a binary unmeasured confounder that were twice as prevalent conditional on $X = c_2$ versus on $X = c_1$ and such that the mean of Y were twofold higher when $U = 1$ versus when $U = 0$ conditional on either $X = c_2$ or $X = c_1$ (all conditional on the measured covariates, $\mathbf{Z}$).

Simplified forms with additional assumptions

Equation 1 can be used to compute an E-value analog when one has access to raw data or published estimates for regression coefficients as well as the marginal expectations of the measured confounders. When conducting sensitivity analyses for papers that report only $\widehat{\beta}_X$ and the marginal sample mean of Y, one can still compute an E-value under the additional assumptions that:

1. X and Y are bivariate normal (marginally on $\mathbf{Z}$).

2. The regression model does not include interactions of X with $\mathbf{Z}$ (or any such interactions are of negligible magnitude).

Then, set $c_1 = E[X]$ and consider the contrast $X = E[X] + \Delta$ versus $X = E[X]$ (choices of Δ are discussed below). Then, the ACE associated with an increase in X of Δ units for a population in which the sample covariate distribution conditional on $X = E[X] + \Delta$ and the sample covariate distribution conditional on $X = E[X]$ each have prevalence 50% satisfies:

$$ACE \geq \left(\widehat{E}\Big[Y \mid X = E[X] + \Delta\Big] - \widehat{E}\Big[Y \mid X = E[X]\Big] \cdot B \right) \times \left(\frac{1+B}{2B} \right)$$

$$= \left(\widehat{E}\Big[Y \mid X = E[X] + \Delta\Big] - \left(\widehat{E}[Y] + \underbrace{\rho_{XY} \frac{\sigma_Y}{\sigma_X} (E[X] - E[X])}_{=0} \right) \cdot B \right) \times \left(\frac{1+B}{2B} \right)$$

$$= \left(\widehat{E}\Big[Y \mid X = E[X] + \Delta\Big] - \widehat{E}[Y] \cdot B \right) \times \left(\frac{1+B}{2B} \right)$$

$$= \left(\widehat{E}[Y] + \Delta \widehat{\beta}_X - \widehat{E}[Y] \cdot B \right) \times \left(\frac{1+B}{2B} \right) \qquad \text{(no } X\text{-}\mathbf{Z} \text{ interactions)}$$

$$= \left((1-B) \widehat{E}[Y] + \Delta \widehat{\beta}_X \right) \times \left(\frac{1+B}{2B} \right)$$

where the second line uses a well-known property regarding conditional expectations in a bivariate normal distribution (in which ρ_{XY} is the correlation between X and Y and σ_Y and σ_X are marginal standard deviations). Equivalently, under the same two assumptions, one can use the estimated correlation between X and Y (denoted r_{XY}), partialling out any measured covariates $\mathbf{Z}$, along with the estimated marginal standard deviations of X and Y ($\widehat{\sigma}_Y$ and $\widehat{\sigma}_X$):

$$ACE \geq \left((1-B) \widehat{E}[Y] + \Delta r_{XY} \frac{\widehat{\sigma}_Y}{\widehat{\sigma}_X} \right) \times \left(\frac{1+B}{2B} \right)$$

Again, the E-value can then be computed through a numerical search for the B such that the left-hand side is 0.

In general, we would recommend using $\Delta = 1$ in order to assess the E-value for the effect size corresponding directly to the regression coefficient, which represents a 1-unit contrast in X. However, if the units of X are very fine-grained (e.g., if X is blood pressure in mmHg), then a 1-unit increase may not be considered clinically meaningful, and a different choice of Δ may be used (e.g., $\Delta = 10$ to represent an increase in blood pressure of 10 mmHg), which is equivalent to rescaling the regression coefficient. It is imperative to report the choice of Δ if it is not taken to be 1, since it directly impacts the size and interpretation of the E-value analog.

References

Ding, Peng, and Tyler J VanderWeele. 2016. "Sensitivity Analysis Without Assumptions." *Epidemiology* 27 (3). Wolters Kluwer Health: 368.

VanderWeele, Tyler J, and Peng Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine* 167 (4). Am Coll Physicians: 268–74.