WEB APPENDIX

Distribution of Lognormal Biomarkers with Error

Let $X \sim logNormal$, implying that $\log(X) \sim N(\mu_X, {\sigma_X}^2)$. The mean and variance of X in terms are μ_X and , ${\sigma_X}^2$ are $\mathrm{E}[X] = e^{\mu_X + 0.5\sigma^2{}_X}$ and $\mathrm{Var}[X] = (e^{\sigma^2{}_X} - 1)e^{2\mu_X + {\sigma_X}^2}$, respectivley.

Suppose $X^* = X + \varepsilon$, where $\varepsilon \sim N(0, {\sigma_\varepsilon}^2)$ is the measurement error and is independent of X. The distribution of X^* is the convolution of a lognormal and normal random variables without a closed form but with the following mean and variance

$$\begin{aligned}\mathrm{E}[X^*] &= \mathrm{E}[X] + \mathrm{E}[\varepsilon] \\ &= e^{\mu_X + 0.5} + {\sigma_X}^2\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}[X^*] &= \mathrm{Var}[X] + \mathrm{Var}[\varepsilon] \\ &= (e^{\sigma^2{}_X} - 1)e^{2\mu_X + {\sigma_X}^2} + {\sigma_\varepsilon}^2\end{aligned}$$

, respectively.

Assuming the availability of the mean and variance from the measurements X^* and the measurement error ε, we can solve for the parameters of interest μ_X and ${\sigma_X}^2$. Let R be the ratio or proportion of total variability attributable to the actual biomarker

$$R = \frac{\mathrm{Var}[X]}{\mathrm{Var}[X^*]} = \frac{\mathrm{Var}[X]}{\mathrm{Var}[X] + \mathrm{Var}[\varepsilon]}$$

then,

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{Var}[X^*] - {\sigma_\varepsilon}^2 \\ (e^{{\sigma_X}^2} - 1)e^{2\mu_X + {\sigma_X}^2} &= \mathrm{Var}[X^*] - {\sigma_\varepsilon}^2 \\ (e^{{\sigma_X}^2} - 1)\,\mathrm{E}[X^*]^2 &= \mathrm{Var}[X^*] - {\sigma_\varepsilon}^2 \\ {\sigma_X}^2 &= log\left[\frac{(\mathrm{Var}[X^*] - {\sigma_\varepsilon}^2)}{\mathrm{E}[X^*]^2} + 1\right] \\ {\sigma_X}^2 &= log\left[\frac{(R \cdot \mathrm{Var}[X^*])}{\mathrm{E}[X^*]^2} + 1\right]\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}[X^*] &= \mathrm{E}[X] \\ \mathrm{E}[X^*] &= e^{\mu_X + 0.5{\sigma_X}^2} \\ \mu_X &= log[\mathrm{E}[X^*]] - 0.5{\sigma_X}^2\end{aligned}$$

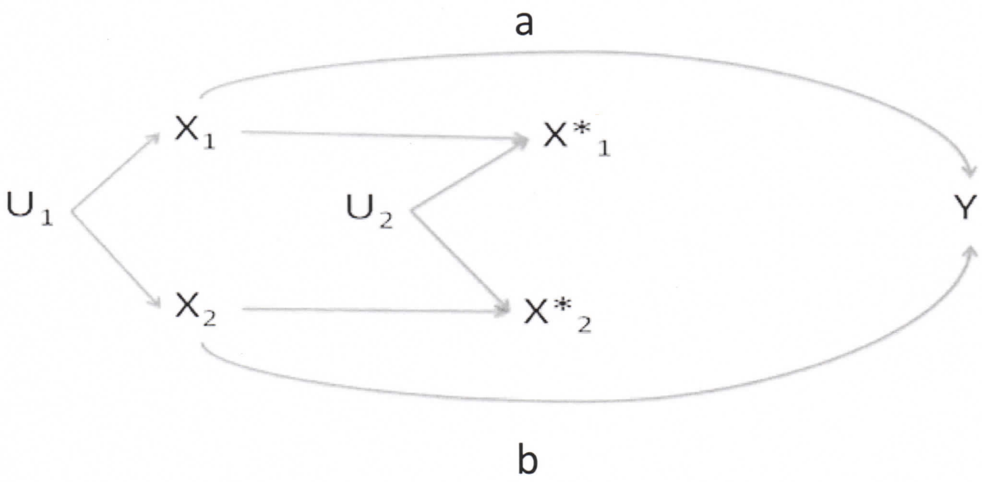
Specifically the cadmium and lead exposures here reasonably follow lognormal distributions and can be characterized by component wise exponentiation multivariate normal distribution,

$$\vec{U} \sim N\left(\vec{\mu}_U = \begin{pmatrix}\mu_{logcd} \\ \mu_{logpb}\end{pmatrix}, \Sigma_U = \begin{bmatrix}\sigma^2_{logcd} & \rho_X\sigma_{logcd}\sigma_{logpb} \\ \rho_X\sigma_{logcd}\sigma_{logpb} & \sigma^2_{logpb}\end{bmatrix}\right).$$

Parameter values in our simulations were chosen based on observed levels in the BioCycle Study with consideration to varying levels of measurement error, $\mu_{logcd} = \log(0.34) - 0.5 \times \sigma^2_{logcd}$, $\sigma^2_{logcd} = \log\left(R \times \frac{0.05}{0.34^2} + 1\right)$ and $\mu_{logpb} = \log(1.02) - 0.5 \times \sigma^2_{logpb}$, $\sigma^2_{logpb} = \log\left(R \times \frac{0.40}{1.02^2} + 1\right)$. Next, we constructed cadmium and lead exposures measured with mean 0 and multivariate normally distributed error, such that $\vec{X}^* = \vec{X} + \vec{\varepsilon}$, where $\vec{\varepsilon} \sim N\left(\vec{\mu}_\varepsilon = \begin{pmatrix}0 \\ 0\end{pmatrix}, \Sigma_\varepsilon = \begin{bmatrix}\sigma^2_{\varepsilon cd} & \rho_\varepsilon\sigma_{\varepsilon cd}\sigma_{\varepsilon pb} \\ \rho_\varepsilon\sigma_{\varepsilon cd}\sigma_{\varepsilon pb} & \sigma^2_{\varepsilon pb}\end{bmatrix}\right)$,

$\sigma^2_{\varepsilon cd} = (1 - R) * 0.05$ and $\sigma^2_{\varepsilon pb} = (1 - R) * 0.40$. This method was used for all the scenarios of correlations and R.

Web Figure 1. Directed acyclic graph of the association between cadmium (X_1), lead (X_2), measured cadmium (X_1^*), measured lead (X_2^*), true correlation between biomarkers ($U_1 = \rho_x$), error correlation ($U_2 = \rho_\varepsilon$) and anovulation (Y).



Linear Approximation to Logistic Bias

Since

$$\beta = cov(X^*)^{-1} cov(X^*, Y)$$

$$cov(X_1^*, Y) = cov(X_1 + \varepsilon_1, Y) = cov(X_1, Y) = a$$

$$cov(X_2^*, Y) = cov(X_2 + \varepsilon_2, Y) = cov(X_2, Y) = b,$$

we have $\beta = cov(X^*)^{-1} \binom{a}{b}$.

Since

$$\beta_{true} = cov(X)^{-1} cov(X, Y)$$

$$cov(X) = cov(X^* - \varepsilon) = cov(X^*) + cov(\varepsilon),$$

we have $\beta_{true} = [cov(X^*) + cov(\varepsilon)]^{-1} \binom{a}{b}$.

$$Bias = \beta - \beta_{true}$$

$$= \left[\left(cov(X) + cov(\varepsilon)\right)^{-1} - \left(cov(X)\right)^{-1}\right] cov(X,Y)$$

$$= \left\{\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} - \begin{pmatrix} \sigma_1^2 + \sigma_{\varepsilon 1}^2 & \rho\sigma_1\sigma_2 + \rho_\varepsilon\sigma_{\varepsilon 1}\sigma_{\varepsilon 2} \\ \rho\sigma_1\sigma_2 + \rho_\varepsilon\sigma_{\varepsilon 1}\sigma_{\varepsilon 2} & \sigma_2^2 + \sigma_{\varepsilon 2}^2 \end{pmatrix}^{-1}\right\} \begin{pmatrix} a \\ b \end{pmatrix}.$$

Web Figure 2. Levels of bias for main effects models for closed form solution under varying levels of correlation between biomarkers (ρ_x) and error correlation (ρ_ε) for the association between two continuous exposure biomarkers and continuous outcome under high levels of measurement error (R=0.50) for main effects of OR=0.75 (circle), 1.00 (triangle), 1.25 (+), 1.75 (x). Level of ρ_x varies along the top panel. Points between −0.6 and 0.6 are noted within each figure and refer to levels of ρ_ε.

Web Table 1. Relative percent bias (mean squared error) for main effects models of the association between cadmium (Cd), lead (Pb) and anovulation, with moderate levels of measurement error (R=0.80) for OR 0.75 and OR 1.75 for *n*=200, for varying levels of correlation between biomarkers and error correlation

	ρ_ε −0.6		**ρ_ε 0.0**		**ρ_ε 0.6**	
	Cd	**Pb**	**Cd**	**Pb**	**Cd**	**Pb**
OR = 0.75						
ρ_x −0.6	−16.17 (0.60)	−16.23 (0.08)	−47.76 (0.53)	−26.35 (0.07)	−79.34 (0.58)	−31.52 (0.08)
ρ_x −0.2	10.04 (0.47)	−11.77 (0.06)	−23.14 (0.46)	−19.46 (0.06)	−55.64 (0.47)	−21.90 (0.07)
ρ_x 0.0	11.80 (0.45)	−11.00 (0.06)	−6.87 (0.45)	−17.44 (0.07)	−46.23 (0.49)	−17.87 (0.06)
ρ_x 0.2	16.85 (0.45)	−12.48 (0.06)	−12.73 (0.44)	−16.38 (0.06)	43.29 (0.50)	−13.60 (0.07)
ρ_x 0.6	43.12 (0.52)	−14.60 (0.07)	13.68 (0.59)	−20.39 (0.08)	−18.99 (0.67)	−18.28 (0.09)
OR = 1.75						
ρ_x −0.6	−10.33 (0.60)	−18.56 (0.11)	−43.36 (0.65)	−26.42 (0.11)	−73.42 (0.77)	−31.55 (0.13)
ρ_x −0.2	8.19 (0.49)	−14.41 (0.08)	−23.83 (0.50)	−22.28 (0.10)	−48.19 (0.59)	−24.28 (0.10)
ρ_x 0.0	13.34 (0.49)	−14.69 (0.08)	−15.30 (0.46)	−19.28 (0.09)	−43.83 (0.57)	−23.72 (0.10)
ρ_x 0.2	17.84 (0.48)	−14.13 (0.08)	−16.18 (0.48)	−19.84 (0.08)	−36.53 (0.56)	−19.81 (0.09)
ρ_x 0.6	39.67 (0.61)	−19.68 (0.09)	6.39 (0.60)	−20.55 (0.10)	−19.57 (0.74)	−23.14 (0.13)

Web Table 2. Relative percent bias (mean squared error) for models of an interaction (CdxPb) between cadmium (Cd), lead (Pb) and anovulation, with moderate levels of measurement error ($R = 0.80$) for OR 0.75 and OR 1.75 for $n = 200$, for varying levels of correlation between biomarkers and error correlation

	ρ_ε −0.6			**ρ_ε 0.0**			**ρ_ε 0.6**		
	Cd	**Pb**	**CdxPb**	**Cd**	**Pb**	**CdxPb**	**Cd**	**Pb**	**CdxPb**
OR = 0.75									
ρ_x −0.6	58.05 (1.52)	7.46 (0.18)	−81.71 (1.94)	47.21 (1.65)	5.28 (0.19)	−119.88(2.29)	−43.29 (1.72)	8.22 (0.20)	−152.28(2.66)
ρ_x −0.2	50.30 (1.82)	3.47 (0.21)	−42.98 (1.68)	23.28 (1.87)	−2.58 (0.22)	−57.37 (1.76)	−0.70 (1.94)	−7.28 (0.23)	−71.89 (1.92)
ρ_x 0.0	42.63 (1.98)	−1.83 (0.23)	−25.75 (1.61)	2.64 (2.06)	−10.12 (0.23)	−35.83 (1.68)	−32.77 (2.06)	−18.43 (0.24)	−35.66 (1.65)
ρ_x 0.2	35.25 (2.22)	−8.95 (0.26)	−7.33 (1.64)	−17.87 (2.21)	−21.37 (0.26)	−5.71 (1.65)	−70.03 (2.21)	−32.51 (0.27)	2.74 (1.56)
ρ_x 0.6	−26.24 (2.58)	−38.71 (0.32)	41.83 (1.64)	−81.61 (2.50)	−52.57 (0.33)	−49.92 (1.53)	−139.49 (2.71)	−58.00 (0.33)	56.44 (1.46)
OR = 1.75									
ρ_x −0.6	51.96 (1.72)	5.21 (0.20)	−84.04 (2.80)	35.65 (1.78)	1.99 (0.21)	−118.26 (3.89)	23.62 (1.82)	1.26 (0.23)	−141.12 (4.83)
ρ_x −0.2	68.08 (2.28)	6.15 (0.24)	−56.50 (2.32)	33.79 (2.11)	−1.83 (0.25)	−70.09 (2.63)	10.77 (2.04)	−4.14 (0.26)	84.31 (2.91)
ρ_x 0.0	56.26 (2.43)	−0.60 (0.26)	−37.73 (2.17)	16.45 (2.24)	−10.00 (0.27)	−46.58 (2.22)	−17.29 (2.13)	−15.36 (0.28)	53.18 (2.21)
ρ_x 0.2	49.05 (2.61)	−5.22 (0.30)	−23.63 (2.10)	−4.21 (2.28)	−18.51 (0.30)	−22.97 (1.96)	−44.51(2.44)	−26.75 (0.32)	22.96 (1.90)
ρ_x 0.6	17.94 (3.01)	−27.14 (0.39)	12.34 (2.25)	−56.11 (3.17)	−44.03 (0.47)	29.24 (2.31)	−107.94 (3.58)	−51.53 (0.50)	33.50 (2.11)