Bayesian and Frequentist Corrections for Covariate Measurement Error

TJ Weaver

University of Rochester

October 23, 2024

Overview

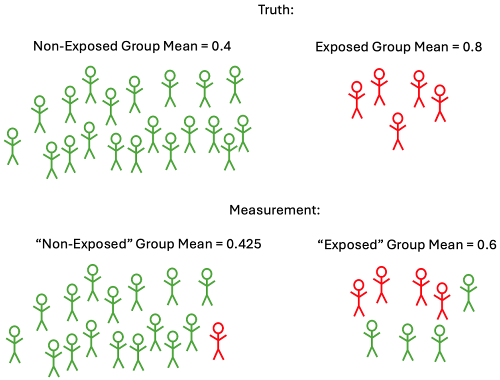
1. Measurement Error
2. Regression Calibration
3. Bayesian Correction
4. Simulation Comparison
5. Example
6. R & Conclusion

What is Measurement Error (and why do we care)?

- Measurement error arises when we have error in the predictors, leading to biased inferences
- In a controlled experiment, measurement error is often not a significant factor, but because epidemiology is observational the investigator often has to use an error-laden measurement of the exposure to a toxin instead of the true exposure
- In a typical linear regression, we have the model $Y = X\beta + \epsilon$, but if the measurement contains error we only have access to W (the erroneous measurement of X) and naively plugging in W for X can obscure the true relationship between X and Y

Categorical Bias Toward the Null

- To better demonstrate how this works, imagine that X is binary, and represents exposure a toxin, and Y is a negative health outcome where higher is worse, and X is measured with 20 % error



Models of measurement error

Typically, we deal with the situation in which the exposure is continuous (not binary), and we assume one of two kinds of error

- Classical Errors: The errors are assumed to be distributed around the true X's, that is $E[W|X] = X$.
- Berkson Errors: The true X's are assumed to be distributed around the error's, $E[X|W] = W$. For example: let's say a machine delivers true doses of a drug "X" distributed around the dial setting of the machine that dictates how much of the drug the machine is supposed to deliver "W". We can only observe W, not the true X, and X is distributed around W

Throughout the rest of this talk we use the classical error model, as assumed by the Bartlett-Keogh paper, and as used in the paper's simulations to be discussed later in the talk.

Continuous Bias Towards the Null

Measurement Error in continuous covariates also biases the relationship between X and Y towards the Null; shown here are 500 observations from the following simulation:

$Y = 1 + 3X + N(0,2),\ W|X \sim N(X,1), X \sim N(0,1)$

Continuous Bias Towards the Null

Aside: In the special case illustrated above, $X \sim N(\mu_X, \sigma_X^2)$, $W|X \sim N(\mu_W, \sigma_W^2)$, the expectation $E[X|W = w] = \frac{\frac{w}{\sigma_W^2} + \frac{\mu_X}{\sigma_X^2}}{\frac{1}{\sigma_W^2} + \frac{1}{\sigma_X^2}} = cw + (1 - c)\mu$ where $c = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}$. While the same linear relationship between X and Y still holds in this special case, the shape of the relationship between X and Y can be affected by the measurement error.

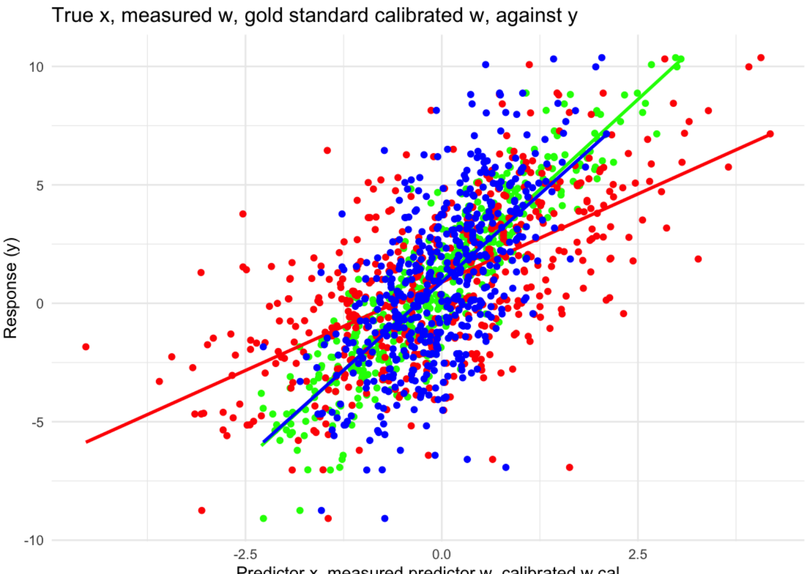
Overview

- The regression calibration approach is one of the simplest ways to address measurement error. Assuming the classical error model $E[W|X] = X$, we simply replace W with an estimate of $E[W|X]$. But how can we estimate$E[W|X]$ without X?
- Usually, a substudy is done so that for a small proportion of the population we have a measurement with error and a measurement without error.
- That is, for a subpopulation of our study we have the true X, often through a ”Gold Standard” measure of the covariate of interest that is too expensive to apply for the entire study.

Regression Calibration

- One easy and generalizable method to estimate $E[W|X]$ is to simply, for the substudy where x is known, regress x on w, fit a model $\hat{X} = E[X|W = w] = \alpha_1 w$ (assuming W, Y are centered), and then plug all the error-laden measurements W into the model, and use the results to fit the model $Y = \beta_0 + \beta_1 \hat{X} + \epsilon$
- This method generalizes nicely, for example in a polynomial model one could simply replace x by $E[X|W = w]$, replace x^2 by $E[X^2|W = w]$, etcetera. The results of this approach on the previous example, with a substudy of size 50:

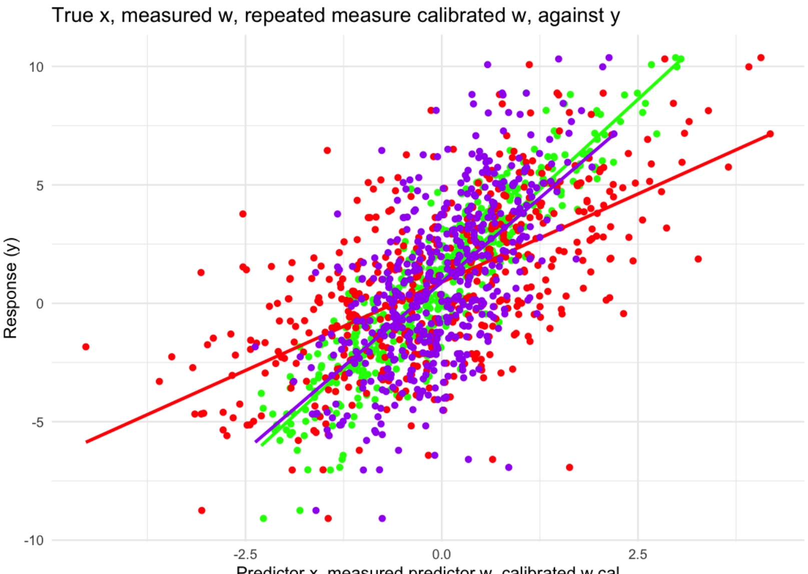
Regression Calibration with "Gold Standard"



Repeated Measures

- When it is not possible to measure a "Gold Standard" an alternative approach is to, for a substudy, repeat the error prone measurement of X, so that for some subpopulation of the study we have one or several more measurements from which to estimate the effect of the error. This is what was done in the Bartlett -Keogh simulations.
- The simplest version of RC for repeated measures is simply repeating the same method as before but replacing the "Gold standard" X's with the error laden repeated measurements (second repetitions).
- In other words, for replicated individuals i, we regress W_{i2} on W_{i1}, the results of this type of calibration on the simulation with 100 of the 500 measures having a second repetition:

Repeated Measures



Efficient Repeated Measures

- When some additional assumptions are made, such as that $X_i|Z_i \sim N(\gamma_0 + \gamma_Z^T Z_i, \sigma^2_{X|Z})$, and the measurement errors are normally distributed, etc. more efficient versions of Regression Calibration can be used.
- So far we have only talked about the linear model in which RC gives consistent parameter estimates. Regression calibration approaches in the logistic and cox models are also approximately consistent under certain assumptions.

Regression Calibration for Logistic and Cox

In order to have consistency for RC in logistic, we assume:

- The outcome is rare OR $\beta_X^2 Var(X_i|W_i)$ is small
- $X_i|W_i$ is normal
- $Var(X_i|W_i)$ is 'small'

In order to have consistency for RC in cox regression, we assume:

- The event rate is low OR measurement error variance is small

Model Specification

For the simulations the paper conducts, with W_{i1} as the measurements with error, W_{i2} as the repeated measurements with error in the replication substudy, and Z_i as the covariates, a bayesian model is specified to determine the relationship between Y and X via posterior inference on β

The posterior inferences of this model are then compared to the results of the regression calibration approach for making inferences about β for Linear, Logistic, and Cox regression, with differing levels of measurement error and different levels of correlation between X and Y

Joint Model

- First, a joint model is specified for $(Y_i, X_i, W_{i1}, W_{i2}|Z_i)$, by conditioning on Z there is no need to model its distribution
- The joint model can be decomposed as $f(Y_i|X_i, Z_i, \beta, \nu)f(W_i|X_i, \sigma_U^2)f(X_i|Z_i, \gamma)$
- U is the measurement error, the first component is the outcome model, the second component is the measurement model, and W_{ij} are assumed to follow a classical error model, the final component specifies a model for the unobserved covariate X_i conditional on Z_i with a default choice of the normal linear regression model
- ν denotes additional parameters, for example the baseline hazard function in cox regression

Prior Specification

In a bayesian approach, some prior distribution must be specified for the model parameters, in this analysis commonly used ”reference priors” which have minimal impact on inferences are assumed:

- β and γ have diffuse normal priors centered at 0
- Gamma(0.5, 0.5) is used as the prior for the precision (reciprocal of variance) parameters
- For the Cox model, the baseline hazard process is assumed to be independent of the other priors and utilizes a Gamma process prior $H_0(t) = GP(cH_0*, c)$, where $H_0 * (t)$ is a prior guess at the mean and c is confidence in that guess, with small c corresponding to a diffuse prior

Posterior Simulation

- Posteriors draws are simulated with MCMC
- Under certain regularity conditions, as the sample size tends to infinity the choice of prior has no impact on the posterior distribution, since it is dominated by the likelihood function
- As a result the bayesian posterior mean estimator is consistent, asymptotically normal, and efficient
- Model coded in JAGS with R

Advantages of the Bayesian Approach

There are several key advantages of bayesian models for measurement error, which include

- Dealing with measurement error with a ML based approach often involves intractable integrals, which is overcome in the bayesian setting through MCMC
- Software to fit models which allow for measurement error is limited, while software for MCMC sampling is very flexible and can be used to handle measurement error by building it into the model
- External information about the measurement process can be incorporated easily in the bayesian setting
- Bayesian modeling can be used to handle both measurement error and missing data

Simulation Studies

Three types of simulations were used to compare RC to Bayesian methods

- Linear: $\beta_0 = 0, \beta_X = \beta_Z = 1$, σ_X^2 chosen st $R^2 = 0.1, 0.5, 0.9$, $\frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2} = 0.5, 0.7, 0.9$
- Logistic: Intercept chosen such that $P(Y = 1) = 0.2$, log odds ratios $\beta_X = 0.1, 0.5, 2$ chosen to represent small, moderate, and large effects of X_i, $\frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2} = 0.5, 0.7, 0.9$
- Cox: Event times generated with a Weibull hazard model $h(t|X_i, Z_i) = \kappa\lambda t^{\kappa-1} e^{\beta_X X_i + \beta_Z Z_i}$, such that 10 % of individuals had an event time before the end of "follow up" at time 10, log hazard ratios $\beta_X = 0.1, 0.5, 2$ and $\beta_X = \beta_Z$
- Simulations run 1000 times (except for Cox, which was only run 100 times due to computational burden)

Models

The data generation models explicitly were:

- Linear: $Y_i = \beta_0 + \beta_X X_i + \beta_Z Z_i + \epsilon$
- Logistic: $logit(prob(Y_i = 1)) = \beta_0 + \beta_X X_i + \beta_Z Z_i$
- Cox: $h(t|X_i, Z_i) = \kappa\lambda t^{\kappa-1} e^{\beta_X X_i + \beta_Z Z_i}$
- Illustrated Example (later in presentation):

$$h(t) = rt^{r-1} exp(\beta_0 + \beta_1 \text{sbp}_i + \beta_2 \text{sex}_i + \beta_3 \text{age}_i + \beta_4 \text{smoker}_i + \beta_5 \text{diabetes}_i)$$

Simulation Results for Linear Regression

Reliability	R^2	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	1.03 (0.28)	1.17 (0.39)	0.94
0.5	0.5	1.03 (0.19)	1.15 (0.27)	0.92
0.5	0.9	1.03 (0.17)	0.98 (0.07)	0.98
0.7	0.1	1.01 (0.20)	1.05 (0.22)	0.95
0.7	0.5	1.01 (0.09)	1.04 (0.10)	0.94
0.7	0.9	1.00 (0.07)	1.01 (0.05)	0.97
0.9	0.1	1.00 (0.16)	1.02 (0.16)	0.95
0.9	0.5	1.00 (0.06)	1.01 (0.06)	0.94
0.9	0.9	1.00 (0.03)	1.01 (0.03)	0.94

Simulation Results for Logistic Regression

Reliability	β_X	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	0.10 (0.12)	0.12 (0.14)	0.94
0.5	0.5	0.51 (0.15)	0.58 (0.19)	0.94
0.5	2	1.64 (0.31)	1.94 (0.32)	0.97
0.7	0.1	0.10 (0.10)	0.11 (0.10)	0.95
0.7	0.5	0.50 (0.11)	0.52 (0.12)	0.94
0.7	2	1.74 (0.20)	2.01 (0.27)	0.97
0.9	0.1	0.10 (0.09)	0.10 (0.09)	0.94
0.9	0.5	0.50 (0.10)	0.51 (0.10)	0.95
0.9	2	1.91 (0.17)	2.01 (0.20)	0.96

Simulation Results for Cox Regression

Reliability	β_X	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	0.10 (0.09)	0.10 (0.09)	0.98
0.5	0.5	0.49 (0.11)	0.48 (0.11)	0.94
0.5	2	1.49 (0.15)	1.92 (0.20)	0.92
0.7	0.1	0.11 (0.09)	0.11 (0.09)	0.98
0.7	0.5	0.49 (0.11)	0.48 (0.11)	0.93
0.7	2	1.67 (0.16)	1.98 (0.18)	0.97
0.9	0.1	0.11 (0.10)	0.11 (0.10)	0.96
0.9	0.5	0.51 (0.10)	0.50 (0.10)	0.96
0.9	2	1.84 (0.15)	1.96 (0.15)	0.95

Modeling Cardiovascular Disease with Bayesian Cox and RC Cox

- Given that the greatest advantage of the Bayesian approach seemed to be in Cox Regression, and to illustrate its flexibility, the approach was applied to the NHANES III data, a survey conducted in the US between '88 and '94 in 33,994 individuals two months and older
- Death due to cardiovascular disease was modeled as the event of interest with blood pressure (sbp, subject to measurement error), sex, age, smoking status, and diabetes as covariates
- To improve fitting speed, inference for a Weibull regression model of the hazard was used when evaluating the measurement error model, so that the hazard was modeled as $h(t) = rt^{r-1}exp(\beta X)$, r given a diffuse prior

Naive Analyses

- Fitting the Naive Weibull and Naive Cox revealed that the Weibull assumption was reasonable since the estimates of the log hazard ratio were close to one another with and without the Weibull assumption
- The test of the Schoenfeld residuals gave p = 0.08, insufficient evidence to reject the proportional hazards assumption
- Through fitting a logistic regression model for misssingness of the smoking variable, there was evidence smoking was more likely to be missing for females, older individuals, diabetics, and individuals with longer follow up times, indicating the complete case analysis may be biased by the exclusion of missing variables

Measurement Error Analyses

- RC was fit in two stages, a LMEM with a random effect for individual and fixed effects for sex, age, smoking, and diabetes, and a fixed effect for systematic shift in mean between first and second exams, used to model the true SBP at exam one, which was then used as a covariate in the weibull regression, where CIs were obtained via bootstrap
- The Bayesian analysis was fit with diffuse normal priors on regression coefficeints and Ga(0.5, 0.5) on the precisions, with sbp assumed to follow classical error, $sbp_{ij} = sbp_i + U_{ij}, j \in 1, 2$ where 1 and 2 are the repeated measures, $U_{i1} \sim N(0, \sigma_U^2)$
- To accommodate missingness under MAR, a model was assumed for the distribution of smoking: $logit P(smoker_i = 1) = \alpha_0 + \alpha_1 sex_i + \alpha_2 age_i + \alpha_3 diabetes_i$, with independent diffuse normal priors for all regression coefficients

Missingness

- The inclusion of the partially missing observations (which added 3,852 individuals to the 2,667 complete ones) improves efficiency and reduces the size of the credible intervals
- The authors considered using MI in conjunction with RC, but implementation was not straightforward because valid within imputation estimates are required, necessitating the use of bootstrap or estimating equation variance estimators

Results

Table 5. Log hazard ratios estimates and 95% confidence/credible intervals for the NHANES III data.

Covariate	Naive CCA	Naive Bayes	RC CCA	Bayes adj. CCA	Bayes adj. full
SBP (per 20 mmHg)	0.085 (0.014, 0.157)	0.086 (0.015, 0.160)	0.115 (0.014, 0.221)	0.114 (0.017, 0.211)	0.122 (0.059, 0.186)
Male	0.49 (0.30, 0.67)	0.49 (0.32, 0.67)	0.49 (0.32, 0.68)	0.49 (0.31, 0.69)	0.46 (0.36, 0.57)
Age (per 10 years)	0.88 (0.77, 0.99)	0.87 (0.76, 0.99)	0.87 (0.76, 0.99)	0.87 (0.75, 0.98)	1.01 (0.94, 1.09)
Smoker	0.26 (0.07, 0.46)	0.25 (0.06, 0.45)	0.26 (0.07, 0.45)	0.26 (0.06, 0.46)	0.24 (0.07, 0.41)
Diabetes	0.50 (0.29, 0.72)	0.50 (0.28, 0.72)	0.50 (0.28, 0.72)	0.50 (0.27, 0.71)	0.68 (0.55, 0.81)

CCA: complete case analysis performed using 2667 individuals, full analysis performed using 6519 individuals; RC: regression calibration; naïve: ignoring measurement error; adj.: adjusting for measurement error in SBP; SBP: systolic blood pressure.

Other interesting findings

- For those interested, two RMD's of my own code and a link to the github repository of the analysis performed by Bartlett and Keogh on the NHANES data have been sent out to the class, which we can cover if there is time
- The first creates the graphs shown earlier in the presentation
- The second is a simplistic simulation that demonstrates a simple bayesian and a simple RC correction for measurement error

Conclusion

Bayesian corrections for measurement error exhibit comparable performance to RC in simulations and may be useful for frequentist as well bayesian statisticians, given that they exhibit good frequentist large sample properties. Bayesian is well suited for measurement error due to the ease of dealing with both measurement error and missingness simultaneously. In addition, the uncertainty intervals automatically allow for skewness typically found in covariate measurement error adjusted estimators. However, as a fully parametric approach, violations of the distributional assumptions can cause problems, and addressing such problems with more complex models can cause model fits to take tremendously long (eg Cox model). Further research is warranted to develop software that can better implement the bayesian approach.

References

Bartlett, Jonathan W., and Ruth H. Keogh. ”Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration.” *Statistical Methods in Medical Research* 27.6 (2018): 1695-1708.

Thomas, Duncan, Daniel Stram, and James Dwyer. ”Exposure measurement error: influence on exposure-disease relationships and methods of correction.” *Annual review of public health* 14.1 (1993): 69-93.

Questions?

Thanks for your attention!