Merc: A New R Package for Correcting Measurement Error Bias based on Regression Calibration Method

Poster Code: MSR60

Xiaofan Liu[1 2], Xin Zhou[1]

[1] Department of Biostatistics, School of Public Health, Yale University, New Haven, CT, USA

[2]Department of Pharmaceutical and Health Economics, Alfred E. Mann School of Pharmacy, University of Southern California, Los Angeles, CA, USA

USC Mann — Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

USC Schaeffer — Leonard D. Schaeffer Center for Health Policy & Economics

Objective

The "merc" R package implements the **Regression Calibration** method for **reliability study** to correct **measurement error bias** in observational studies which can be applied to **linear regression, logistic regression and cox regression.**

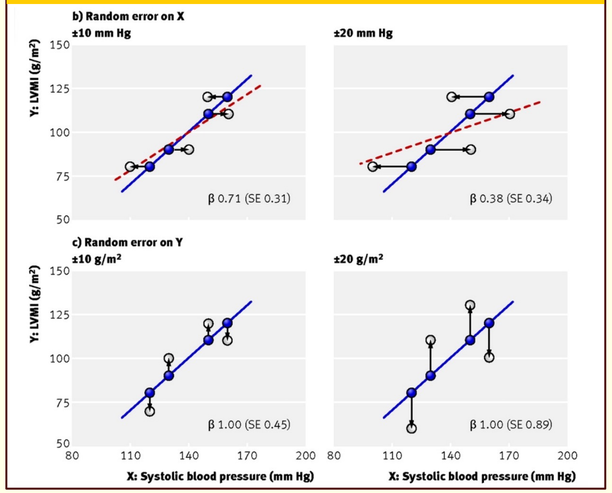
Introduction

- **Random measurement error** in exposure can **cause bias towards the null** in the estimation of regression **coefficients and variance (Attenuation Bias)**
- **Walter Willett, Bernard Rosner, and Donna Spiegelman** at **Harvard T.H. Chan School of Public Health**, developed **Regression Calibration** method to correct measurement error bias.
- There is **very few statistical software** to implement it.

Table1: Available Software for Regression Calibration

Name	Software	Methods
%blinplus	SAS macro	Validation study
%relibpls8	SAS macro	Reliability study
rcal in merror	STATA package	Validation/reliability study
mecor	R package	Only Linear regression

Figure1: Effect of Random Measurement Error on Simple Linear Regression (Y = left ventricular mass index, X = systolic blood pressure)



Methods

Regression Calibration for Reliability Study

- **Reliability study** has been developed to assess the relation between **observed replicates of exposure (subject to random within-person variability)** and **true exposure.**
- **The average of repeated measurements within a short period** of time could be considered **as approximation** for **true exposure** if there is no gold standard for measurement, such as **systolic blood pressure, serum hormones, serum glucose**.

Algorithm:

Step1. In **main study,** obtain the **uncorrected** regression coefficients $\boldsymbol{\beta}'_{(s+t+1)\times 1}$.
Let $\tilde{\mathbf{Z}}$ = (Z, C). Z is a s ×1 vector of mis-measured continuous covariates and C is a t ×1 vector of perfectly measured covariates.

Step2. In **reliability study**, estimate **within-person covariance matrix**. $\boldsymbol{\Sigma}$

$$\Sigma_{s\times s} = \sum_{i=1}^{n_R}\sum_{j=1}^{n_{Ri}} \left(Z_{ij} - \bar{Z}_i\right)\left(Z_{ij} - \bar{Z}_i\right)^T / \sum_{i=1}^{n_R}(n_{Ri} - 1)$$

$$\Sigma_{(s+t)\times(s+t)} = \begin{bmatrix} \Sigma_{s\times s} & 0_{s\times t} \\ 0_{t\times s} & 0_{t\times t} \end{bmatrix}$$

Step3. In **main study**, estimate the **total covariance matrix** $\boldsymbol{\Sigma}_{\bar{\mathbf{z}}}$, and derive an estimate of the **between-person covariance matrix** $\boldsymbol{\Sigma}_X$, **by** $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_{\bar{z}} - \boldsymbol{\Sigma}$

Step4. Calculate **multivariate reliability coefficients** $\widehat{\boldsymbol{R}}$

$$\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{\Sigma}_X} \cdot (\widehat{\boldsymbol{\Sigma}_z})^{-1}$$

Step5. Calculate **corrected coefficients** and estimate **variance** for the corrected coefficients based on **multivariate delta method**

$$\boldsymbol{\beta}'_{corrected} = \boldsymbol{\beta}'_{uncorrected} \times \widehat{\boldsymbol{R}}^{-1}$$

Linear Addictive Measurement Error Model:

$Z_{ij} = X_i + e_{ij}$,

$E(e_{ij}) = 0$, $\boldsymbol{\Sigma}_Z = \boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}$

Z_{ij} is replicate for mis-measured exposure, X_i is true value

Table2: Data Structure of Reliability Study

	Y (Outcome)	Z (Mis-measured)	C (Perfectly measured)
Main Study Dataset			
obs 1	y_1	z_1	c_1
...	...	...	...
obs n	y_n	z_n	c_n
Reliability Study Dataset(e.g., 3 replicates for Z)			
	Z_1	Z_2	Z_3
obs 1	z_{11}	z_{12}	z_{13}
...	...	...	...
obs n	z_{n1}	z_{n2}	z_{n3}

Results

" mercRel" Function

- **"mercRel" is the main function to correct regression coefficients, standard error, and 95 % confidence intervals for reliability study.**
- Detailed description of arguments in the "mercRel" function can be found in the **help documents** in R.

Installation

Install the development version from GitHub with:

```
# install.packages("devtools")
devtools::install_github("vanessaxiaofan/merc")
```

Simulation Studies

We provided detailed illustration of how to use the package for different regression(**linear, logistic, cox**) with simulated datasets in **help documents**. Here, due to limited space, we only take **logistic regression as the example.**

- **Dataset (Main study & Reliability Study)**

The simulated dataset **"mainLogRel"** and **"relibLog"** represent a reliability study with binary outcome using Logistic Regression.

- ❑ **"mainLogRel":** The main study dataset which has 1500 subjects with continuous outomce(Y), exposure measured with error(X) and one confounder measured without error(S).
- ❑ **"relibLog" : T**he reliability study dataset which has 700 subjects each with 4 replicates for the mismeasured exposure (X1, X2, X3, X4).
- ❑ The **prevalence of the outcome Y** is around 0.105 which satisfies the rare outcome assumption.
- ❑ **Intra class correlation(ICC)** for the reliability study is 0.7.

R code

```
# Check help document
R > library(merc)
R > ?mercRel
# Use "mainLogRel" and "relibLog" datasets built in the "merc"
R > data("mainLinearRel", package = "merc")
R > data("relibLog",package = "merc")
# Create weights for covariates (weight =1 in this example)
R > wts <- data.frame(x=1,s=1)
# Apply "mercRel" function to correct the coefficient (OR)
R > mercRel(supplyEstimates=FALSE, relib = relibLog, sur = c("x"), + +
woe = c("s"), weri = c("x","x2","x3","x4"), outcome = c("Y"), rr=4, +
ms=mainLogRel, method = "glm", family = binomial, link = "logit", +
weights=wts)
```

Output: Coefficients (OR) and 95% Confidence Interval (CI) for both Corrected and Uncorrected Model

```
Call:
mercRel(supplyEstimates = FALSE, relib = relibLog, sur = c("x"),  woe = c("s"),
+       outcome = c("Y"), weri = c("x1", "x2", "x3", "x4"), rr = 4, ms = mainLogRel,
+       weights = wts, method = "glm", family = binomial, link = "logit")

Coefficients Uncorrected Model:
  Weights          B   SE(B)   OR(B)  Z Value     Pr(>|Z|) lower 95%CI upper 95%CI
x       1 0.2553199 0.07345 1.29087 3.476105 0.0005087539     1.11779     1.49075
s       1 0.5004175 0.16871 1.64941 2.966140 0.0030156295     1.18500     2.29582

Coefficients Corrected Model:
  Weights          B   SE(B)   OR(B) Z Value     Pr(>|Z|) lower 95%CI(OR) upper 95%CI(OR)
x       1 0.37238 0.10733 1.45118 3.469487 0.000521454          1.17588          1.79095
s       1 0.45034 0.17190 1.56885 2.619779 0.008798678          1.12010          2.19738
```

Conclusion

- Overall, the new R package **"merc"** offers more efficiency and flexibility to implement Regression Calibration method for measurement error correction, compared to existing statistical macro/packages in SAS and STATA.
- The **"merc"** package also contains the function **"mercVal"** for validation study which is written by Wenze Tang, PhD from Harvard T.H. Chan School of Public Health.

Reference

Rosner B, Willett WC, Spiegelman D. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error. Statistics in medicine. 1989;8(9):1051-1069.

Rosner B, Spiegelman D, Willett WC. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: The Case of Multiple Covariates Measured with Error. American journal of epidemiology. 1990;132(4):734-745.

Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. The American Journal of Clinical Nutrition. 1997;65(4):1179S-1186S.

Contact

Email: liuxiaof@usc.edu Phone: (203) 893-5520