Published in final edited form as:

Commun Stat Theory Methods. 2017 ; 46(23): 11604–11611. doi:10.1080/03610926.2016.1275693.

Using repeated measures to correct correlated measurement errors through orthogonal decomposition

Chang Yu^{1,*}, Sanguo Zhang², Christine Friedenreich³, and Charles E. Matthews⁴

¹Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

²School of Mathematical Sciences, Graduate University of Chinese Academy of Sciences, Beijing 100049, China

³Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, 2210 2nd St SW, Calgary, AB T2S 3C3, Canada

⁴Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Bethesda, MD 20892, USA

Abstract

In a physical activity study, the 7-day physical activity log viewed as an alloyed gold standard was used to correct the measurement error in the physical activity questionnaire. Due to correlations between the errors in the two measurements, the usual regression calibration may result to a biased estimate of the calibration factor. We propose a method of removing the correlation through orthogonal decomposition of the errors, then the usual regression calibration can be applied. Simulation studies show that our method can effectively correct the bias.

Keywords

adjusted regression calibration; correlated measurement errors; orthogonal decomposition; physical activity

1 Introduction

Epidemiological studies have established a positive association between higher level of self-reported physical activity (PA) and reduced risk of early mortality, chronic diseases, and certain cancers (Arem et al., 2015; Lee et al., 2012; and Moore et al., 2016). In these studies, error-prone physical activity questionnaires (PAQ) were often used to measure PA since no gold standard measure was available and these PAQs were practical for implementation in a large scale. While these data have given us important insight into the relationship between physical activity and many health outcomes, the PAQ measurements are known to contain a substantial amount of measurement error (Matthews et al., 2012; Tooze et al., 2013). These errors may lead to attenuated estimates of the association between PA and health outcomes, thus a loss of statistical power when testing the hypotheses.

*Contact for corresponding author: chang.yu@vanderbilt.edu, Tel: (615)-322-8422, Fax: (615)-343-4924.

Regression calibration (RC) is often used to correct measurement errors in large cohort studies (Rosner et al. 1989, 1990; Carroll, et al. 2006, Chapter 4). To conduct the RC, we need a gold standard to measure the true exposure in a *validation study* that is usually a sub-study embedded in the *main study*. In the main study, an inexpensive but error-prone measure Q such as a PAQ in PA studies is obtained on all the subjects. Only in the validation study both Q and the true exposure T are obtained. T is measured using the gold standard. Suppose the association between disease outcome D and T follows a logistic model

$$\text{logit} [\Pr(D = 1 | T)] = \alpha + \beta T, \quad (1)$$

and Q has a linear relationship with T

$$T = \lambda_0 + \lambda Q + e, \quad (2)$$

where we assume error e is not correlated with Q and it has $E(e) = 0$ and $\text{Var}(e) = \sigma_e^2$. The RC is a two-step procedure. First, we use Q in place of T in the disease model (1) to analyze the data from the main study. This regression gives us a naive estimate β_N of the association parameter β . In the second step, we calculate the *calibration factor*

$$\lambda_G = \text{Cov}(T, Q) / \text{Var}(Q).$$

This is equivalent to fitting the calibration model (2) to the dataset of the validation study. Then we can obtain the regression-calibrated association estimate

$$\beta_{RC} = \beta_N / \lambda_G. \quad (3)$$

For linear disease models, β_{RC} is unbiased. If the variance of the measurement error tends to zero, the RC estimate is consistent for the logistic model (1) (Stefanski and Carroll, 1985; Carroll and Stefanski, 1990; and Kuha, 1994).

However, in PA studies we do not yet have a gold standard to measure the true PA level T . Instead we only have less-than-perfect *alloyed gold standards*. If a measurement F satisfies

$$F = T + e_F, \quad (4)$$

where error e_F is not correlated with T and it has $E(e_F) = 0$ and $\text{Var}(e_F) = \sigma_{e_F}^2$, measurement

F is called an alloyed gold standard (AGS). In PA studies, the 7-day log is viewed as an AGS. We can rewrite the calibration model (2) as

$$Q = \alpha_Q + \beta_Q T + e_Q, \quad (5)$$

where $E(e_Q) = 0$, $\text{Var}(e_Q) = \sigma_{e_Q}^2$, and $(T, e_Q) = 0$. Note that equations (2) and (5) represent Q and T being projected into different linear spaces. There is no simple relationship $\beta_Q = 1/\lambda$.

If errors e_F and e_Q are not correlated, we can use the AGS F in place of T to calculate the calibration factor

$$\lambda_A = \text{Cov}(F, Q) / \text{Var}(Q).$$

Then λ_A can be used in (3) for the calibration (Spiegelman et al., 1997). The calibrated association estimate has the same properties as if the gold standard T is used. On the other hand, if $\text{Cov}(e_F, e_Q) \neq 0$, Wacholder et al. (1993) showed that the association estimate (3) calibrated using λ_A was no longer consistent for the linear disease model

$$E(D|T) = \alpha + \beta T. \quad (6)$$

Spiegelman et al. (1997) showed that such-calibrated estimate was biased for the logistic model (1). They proposed adding the third instrument in the validation study so that the correlation between e_F and e_Q can be estimated. This estimate forms the basis to correct the correlated errors. The error in the third instrument should not be correlated with either e_F or e_Q .

In this work, we develop a new approach to correcting measurement error in the setting of $\text{Cov}(e_F, e_Q) \neq 0$. Our approach requires repeated measurements of both F and Q in the validation study. We use orthogonal decomposition to obtain a new quantity, F_O , whose error is not correlated with e_Q . Then F_O is used in the calibration as if it is an AGS. We call our method adjusted regression calibration (ARC). We show that the ARC leads to consistent estimate of the association between exposure and disease outcomes for linear models. The estimate is also consistent for the logistic model (1) under the condition the variance of the measurement error tends to zero.

The rest of the manuscript is organized as follows. In Section 2, we develop the ARC method. Simulation results for the linear and logistic models are presented in Section 3. In Section 4, we apply our method to a PA measurement validation study conducted in Canada by Friedenreich and colleagues (Friedenreich et al., 2006). We conclude the manuscript by a discussion. Technical details are given in the Appendix.

2 Adjusted regression calibration

The main step in the RC is to estimate the calibration factor λ . From (2), (4), and (5), we have

$$\lambda_A = \frac{\text{Cov}(F, Q)}{\text{Var}(Q)} = \lambda + \frac{\text{Cov}(e_F, Q)}{\text{Var}(Q)} = \lambda + \frac{\text{Cov}(e_F, e_Q)}{\text{Var}(Q)} = \lambda + \frac{\text{Corr}(e_F, e_Q)\sigma_{e_F}\sigma_{e_Q}}{\text{Var}(Q)}, \quad (7)$$

using the AGS F in place of T . This shows the bias in λ_A depends on the correlation between e_F and e_Q and their variability.

The result (7) suggest we could correct the correlated errors if we can estimate $\text{Cov}(e_F, e_Q)$. Spiegelman et al. (1997) proposed adding the third instrument for the exposure in the validation study to achieve this. We estimate this covariance through repeated measurements of F and Q .

In the validation study, we take the j th ($j = 1, 2, \dots, r_k$) measurement of both the error-prone Q_{kj} and the AGS F_{kj} on subject k ($k = 1, 2, \dots, m$), where m is the number of subjects in the validation study. We have

$$\begin{aligned} F_{kj} &= T_k + e_{F_{kj}}, \\ Q_{kj} &= \alpha_Q + \beta_Q T_k + e_{Q_{kj}}. \end{aligned}$$

We assume $\text{Cov}(T, e_Q) = \text{Cov}(T, e_F) = 0$, and $\text{Cov}(e_{F_{kj}}, e_{Q_{k'j'}}) = 0$ for $k \neq k'$ or $j \neq j'$. In PA studies, we often have $\text{Cov}(e_{F_{kj}}, e_{Q_{kj}}) \neq 0$ since both measurements are self-reported for the same time frame.

Note $\text{Cov}(e_F; e_Q) = \text{Cov}(e_F; Q)$ and let $\mu_Q = E(Q)$, we can then construct

$$F_O = F - \gamma(Q - \mu_Q) = T + [e_F - \gamma(Q - \mu_Q)].$$

By choosing

$$\gamma = \text{Cov}(e_F, Q)/\text{Var}(Q), \quad (8)$$

we can show

$$\text{Cov}(e_F - \gamma(Q - \mu_Q), Q) = 0.$$

From (4) and (7), we observe that $F_O = F - \gamma(Q - \mu_Q)$ can serve as an alloyed gold standard.

Using data collected in the validation study, γ and μ_Q can be estimated by

$$\hat{\gamma} = \sum_{k=1}^m \sum_{j=1}^{r_k} (F_{kj} - F_{k\cdot})(Q_{kj} - Q_{k\cdot}) / (N_m - m) \left[\sum_{k=1}^m \sum_{j=1}^{r_k} (Q_{kj} - \bar{Q})^2 / (N_m - 1) \right]^{-1}, \quad (9)$$

and

$$\hat{\mu}_Q = \bar{Q} = \sum_{k=1}^m \sum_{j=1}^{r_k} Q_{kj} / N_m,$$

where $N_m = \sum_{k=1}^m r_k$, $F_{k\cdot} = \sum_{j=1}^{r_k} F_{kj} / r_k$, $Q_{k\cdot} = \sum_{j=1}^{r_k} Q_{kj} / r_k$. Thus, the RC method of Spiegelman et al. (1997) can be modified by adding a middle step. In this step, we obtain the orthogonal component $F_{O_{kj}}$ of F_{kj} through

$$F_{O_{kj}} = F_{kj} - \hat{\gamma} (Q_{kj} - \bar{Q}). \quad (10)$$

Then we fit calibration model (2) using F_O in place of T . The parameter estimates are

$$\hat{\lambda}_{ARC} = \text{Cov}(F_O, Q) / \text{Var}(Q),$$

and the intercept estimate $\hat{\lambda}_0$. In this fit, we use the average of r_k repeats of $F_{O_{kj}}$ and Q_{kj} for subject k . The calibrated estimates are

$$\beta_{ARC} = \beta_N / \hat{\lambda}_{ARC}, \quad (11)$$

and

$$\alpha_{ARC} = \alpha_N - \beta_{ARC} \hat{\lambda}_0. \quad (12)$$

The naive regression step remains the same.

The following two theorems describe the properties of the ARC estimates. Their proof is in the Appendix.

Theorem 2.1:

For the linear model (6), assuming the model-related conditions are met and

$\liminf_{m \rightarrow \infty} (N_m / m) > 1$, the ARC estimates are consistent.

Theorem 2.2:

For the logistic model (1), assuming the model-related conditions are met, and the conditional distribution of D given T and Q is the same as the conditional distribution of D given T , i.e. $f(D|T, Q) = f(D|T)$, the ARC estimates are consistent if $\liminf_{m \rightarrow \infty} (N_m/m) > 1$ and $\sigma_e^2 \rightarrow 0$.

For estimates (11) and (12) to be consistent, we need condition $\liminf_{m \rightarrow \infty} (N_m/m) > 1$. This requires the average number of measurement repeats per subject cannot be close to 1 in the validation study (Zhang and Chen, 2001). For the logistic model, the condition $f(D|T, Q) = f(D|T)$ means nondifferential measurement error in Q (Carroll et al., 2006, page 36). We need to further assume the variance of the measurement error tends to zero in order for the ARC estimates to be consistent in Theorem 2.2; the same assumption was made in Stefanski and Carroll (1985), Carroll and Stefanski (1990), and Kuha (1994).

3 Simulation studies

We conducted simulation studies to evaluate the ARC estimate in comparison with the naive estimate and the RC estimate. Different strength of the correlation between the measurement errors was simulated for both the linear and logistic models. We examined the bias and the mean squared error (MSE) of the three association estimates.

We first simulated a linear model $D = -1 + T + \epsilon$ with $T \sim N(5, 4)$ and $\epsilon \sim N(0, 1)$. The measurements were set as $Q = 1 + 2T + e_Q$, $F = T + e_F$ where errors were bi-variate normal $(e_Q, e_F)^T \sim N(0, \Sigma)$ with covariance matrix $\Sigma = \begin{pmatrix} 4 & 2\rho \\ 2\rho & 1 \end{pmatrix}$. Each simulation had $n = 500$ subjects in the main study, and we randomly selected 10% ($m = 50$) for the validation study.

In the second simulation, we evaluated the three estimates for a logistic model $\text{logit}[\Pr(D=1|T)] = -5 + 0.5T$ with $T \sim N(2, 4)$. We set the measurements $Q = 5 + 1.5T + e_Q$ and $F = T + e_F$ where errors $(e_Q, e_F)^T \sim N(0, \Sigma)$ with covariance matrix $\Sigma = \begin{pmatrix} 4 & 2\rho \\ 2\rho & 1 \end{pmatrix}$. The main study had a sample size of $n = 1000$, and we randomly selected 5% ($m = 50$) for the validation study. In both simulation studies, each subject had four repeated measurements of Q and F in the validation study. We varied ρ from 0.75 to -0.75 .

The simulation results reported in Table 1 suggest that both the RC and the ARC can correct the bias in the naive estimate when the errors are not correlated. If the correlation between the errors is positive, even though it corrects much of the bias, the RC estimate under-corrects the bias. When the correlation between the errors is negative, the RC estimate over-corrects the attenuation. However, the ARC demonstrates a favorable performance in correcting the bias for both the linear and logistic models in all the simulations.

4 Application to a real dataset

This section reports our application of the ARC to the dataset of a validation study that was conducted between 2002 and 2003 by Friedenreich and colleagues. The study's objective was to examine the reliability and validity of the Past Year Total Physical Activity Questionnaire. Details about the study were described in Friedenreich et al. (2006). The dataset has three repeats of the PAQ and four repeats of the 7-day PA log. The second PAQ did not have the 7-day log measured at the same time. Thus, we included only the first and the third PAQ and their corresponding 7-day logs in this analysis. The total number of subjects was 150, with 73 men and 77 women.

We first evaluated the correlation between the errors. The estimates were $\hat{\rho}_{e_F, e_Q} = 0.10$ with a p -value 0.07 in all the subjects, $\hat{\rho}_{e_F, e_Q} = 0.27$ with a p -value 0.001 for the males, and $\hat{\rho}_{e_F, e_Q} = -0.07$ for the females. The estimate $\hat{\rho}_{e_F, e_Q} = -0.07$ was not statistically significant. These estimates suggest a weak correlation between the errors, notably in the males. Both the RC and the ARC estimates of the calibration factor λ are reported in Table 2. Consistent with the simulation results in Section 3, if the correlation between the errors is ignored, the usual RC would under-correct the attenuation, especially among the males.

5 Discussion

In a setting where the errors in the error-prone measure and the AGS are correlated, we use the orthogonal decomposition of the AGS to remove the correlation. Compared with the method of Spiegelman et al. (1997), our method does not need to add the third instrument in the validation study in order to correct the correlated errors. However, our method needs repeats of both the error-prone and the AGS measurements in the validation study. The RC is essentially a special case of our proposed method when there is no correlation between the errors. Simulation studies indicate that our method can correct the bias in the estimate of the calibration factor for a broad range of correlations between the errors.

Acknowledgment

This work was supported in part by R21CA119073 from NCI (CY, SZ, and CM), R21HL129020 from HLBI, R01FD004778 from FDA, and Vanderbilt CTSA grant UL1 TR000445 from NCATS (CY), and grant No. 10801133 from the National Science Foundation of China (SZ).

Appendix

Proof of Theorem 2.1:

From (8) and (9) under $\liminf_{m \rightarrow \infty} (N_m/m) > 1$, we have $\hat{\gamma} - \gamma = o_p(1)$ (1) due to Zhang and Chen (2001). From the construction of F_O as in (10) and $\bar{Q} - \mu_Q = o_p(1)$, some algebra results to

$$\text{Cov}(F_O, Q)/\text{Var}(Q) = \text{Cov}(F, Q)/\text{Var}(Q) - \gamma + o_p(1).$$

Putting this together with (2) and (4), we have $\text{Cov}(F_Q, Q)/\text{Var}(Q) = \lambda + o_p(1)$. Since the calibration factor is estimated using the corresponding sample variance and covariance, $\hat{\lambda}$ is a consistent estimator of λ and $\hat{\lambda}_0$ is a consistent estimator of λ_0 . Under the linear model (6) and linear calibration model (2), the initial naive regression step gives us $\alpha_N = \alpha + \beta \lambda_0 + o_p(1)$ and $\beta_N = \beta \lambda + o_p(1)$. Then, (11) and (12) lead to the conclusion that β_{ARC} and α_{ARC} are consistent estimators of β and α , respectively.

Proof of Theorem 2.2:

For the logistic model (1) with measurement relationships (2) and (4), due to $f(D | T, Q) = f(D | T)$, the model of D on Q has the form

$$\begin{aligned} P(D = 1 | Q) &= \int P(D = 1 | T) f(T | Q) dT \\ &= \int \frac{\exp(\alpha + \beta T)}{1 + \exp(\alpha + \beta T)} f(T | Q) dT. \end{aligned} \quad (13)$$

Let $\mu(Q) = \lambda_0 + \lambda Q$. Since $T = \lambda_0 + \lambda Q + e = \mu(Q) + e$ and $\sigma_e^2 \rightarrow 0$, an approximation for (13) is obtained by replacing $P(D = 1 | T)$ with its second-order Taylor expansion around $\mu(Q)$ (Kuha, 1994). Then (13) becomes

$$\begin{aligned} P(D = 1 | Q) &= \frac{\exp[\alpha + \beta \mu(Q)]}{1 + \exp[\alpha + \beta \mu(Q)]} \\ &\quad + \frac{\exp[\alpha + \beta \mu(Q)] \{1 - \exp[\alpha + \beta \mu(Q)]\}}{2 \{1 + \exp[\alpha + \beta \mu(Q)]\}^3} \beta^2 \sigma_e^2 + o(\sigma_e^2). \end{aligned}$$

The linear term $[T - \mu(Q)]$ in the above expansion disappears when taking the expectation. Thus, the initial naive regression step gives us $\alpha_N = \alpha + \beta \lambda_0 + o_p(1)$ and $\beta_N = \beta \lambda + o_p(1)$. Since the estimate of the calibration factor does not depend on the outcome data, and in Theorem 2.1 we have proved that λ_b and λ_{b0} are consistent estimators of $\hat{\lambda}$ and $\hat{\lambda}_0$, respectively, Theorem 2.2 follows.

References

- Arem H, Moore SC, Patel A, et al. (2015). Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. *JAMA internal medicine*, 175(6):959–967. [PubMed: 25844730]
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006). *Measurement Error in Nonlinear Models*, second edition, Chapman & Hall, London.
- Carroll RJ, Stefanski LA (1990). Approximate quasi-likelihood estimation in models with surrogate predictors, *JASA*, 85:652–663.
- Friedenreich CM, Courneya KS, et al. (2006). Reliability and validity of the Past Year Total Physical Activity Questionnaire, *American Journal of Epidemiology*, 163:959–970. [PubMed: 16524954]
- Kuha J (1994). Corrections for exposure measurement error in logistic regression models with an application to nutritional data, *Statistics in Medicine*, 13:1135–1148. [PubMed: 8091040]

- Lee IM, Shiroma EJ, Lobelo F, et al. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet*, 380(9838): 219–229.
- Matthews CE, Moore SC, George SM, et al. (2012). Improving self-reports of active and sedentary behaviors in large epidemiologic studies. *Exercise and Sports Science Reviews*, 40(3):118–126.
- Moore SC, Lee I, Weiderpass E, et al. (2016). Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. *JAMA internal medicine*, 176(6):816–825. [PubMed: 27183032]
- Rosner B, Willett WC, Spiegelman D (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error, *Statistics in Medicine*, 8, 1051–1069. [PubMed: 2799131]
- Rosner B, Spiegelman D, Willett WC (1990). 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*, 132, 734–745. [PubMed: 2403114]
- Spiegelman D, Schneeweiss S, McDermott A (1997). Measurement error correction for logistic regression models with an “alloyed gold standard”, *American Journal of Epidemiology*, 145, 184–196. [PubMed: 9006315]
- Stefanski LA, Carroll RJ (1985). Covariate measurement error in logistic regression, *The Annals of Statistics*, 13, 1335–1351.
- Tooze JA, Troiano RP, Carroll RJ, et al. (2013). A measurement error model for physical activity level as measured by a questionnaire with application to the 1999–2006 NHANES questionnaire. *Am J Epidemiol*, 177(11):1199–1208. [PubMed: 23595007]
- Wacholder S, Armstrong B, Hartge P (1993). Validation studies using an alloyed gold standard, *American Journal of Epidemiology*, 137, 1251–1258. [PubMed: 8322765]
- Zhang S, Chen X (2001). Consistency of modified MLE in EV model with replicated observations, *Science in China Series A*, 44, 304–310.

Table 1:

Bias and MSE of the naive, RC, and ARC estimates of the association for simulation studies (I): the linear model (6) with $\beta = 1$ and (II): the logistic model (1) with $\beta = 0.5$ for various correlations between errors e_Q and e_F .

ρ_{e_Q, e_F}	Estimate	(I) Linear model		(II) Logistic model	
		Bias	MSE	Bias	MSE
0.75	β_N	-0.601	0.361	-0.270	0.076
	β_{RC}	-0.157	0.026	-0.099	0.020
	β_{ARC}	0.015	0.005	0.009	0.019
0.25	β_N	-0.599	0.359	-0.270	0.076
	β_{RC}	-0.048	0.006	-0.033	0.014
	β_{ARC}	0.014	0.005	0.008	0.016
0	β_N	-0.600	0.360	-0.270	0.076
	β_{RC}	0.009	0.005	0.007	0.018
	β_{ARC}	0.009	0.005	0.008	0.018
-0.25	β_N	-0.599	0.360	-0.269	0.075
	β_{RC}	0.084	0.016	0.058	0.026
	β_{ARC}	0.011	0.006	0.008	0.018
-0.75	β_N	-0.600	0.361	-0.271	0.077
	β_{RC}	0.272	0.103	0.195	0.090
	β_{ARC}	0.010	0.006	0.004	0.020

Table 2:

The overall and gender-specific estimates of the calibration factor using data from the PAQ and the 7-day PA log validation study of Friedenreich et al. (2006). The confidence interval (CI) was obtained using Bootstrap.

	Overall	Male	Female
$\hat{\lambda}_{ARC}$	0.367	0.312	0.363
95% CI	(0.242, 0.492)	(0.150, 0.474)	(0.155, 0.571)
$\hat{\lambda}_{RC}$	0.401	0.428	0.345
95% CI	(0.276, 0.526)	(0.266, 0.590)	(0.137, 0.553)