Applying Linear Mixed-Effects Models to the Problem of Measurement Error in Epidemiologic Studies

Christopher H. Morrell, Larry J. Brant, Jay D. Pearson, Geert N. Verbeke, and Jerome L. Fleg

1Loyola College in Maryland, Baltimore, Maryland, USA
2Gerontology Research Center, National Institute on Aging, Baltimore, Maryland, USA
3Merck Research Laboratories, West Point, Pennsylvania, USA
4Biostatistical Center, Katholieke Universiteit Leuven, Leuven, Belgium
5NHLBI/DECA, Rockledge 2, Bethesda, Maryland, USA

ABSTRACT

Although studies of the relationship between risk factors measured at...
baseline and a binary outcome are common problems occurring in statistical analyses of epidemiologic studies, little has been written about what constitutes a good measure of the true baseline value of a risk factor. This paper considers different schemes of providing for baseline risk factor data and proposes and compares methods for correcting estimates of relative risks for bias due to measurement error. Repeated measurements taken at baseline appear to allow for a correction in the estimate of risk that is comparable to the correction based on repeated measurements over a number of visits close to baseline. The procedures proposed are regression calibration/imputation type methods that involve estimating the covariates with less error and using these estimates in the standard model for the risk factor analysis. Shrinkage estimates from a linear mixed-effects model for multiple measurements at baseline are proposed as a method for estimating the true level of a risk factor at baseline. A simulation study shows that these shrinkage estimates give estimates of the risk of the outcome which have a smaller mean square error and confidence interval coverage proportions much closer to the nominal level than would be obtained using the observed data. This new imputation method is compared to the measurement error correction method earlier proposed by Rosner et al. (Rosner, B., Spiegelman, D., Willett, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. American Journal of Epidemiology 136:1400–1413). In addition, the standard errors from the new method are compared to a nonparametric bootstrap estimate of the standard errors as well as to the asymptotic standard errors obtained using a method proposed by Carroll and Stefanski (Carroll, R. J., Stefanski, L. A. (1990). Approximate Quasi-likelihood Estimation in Models With Surrogate Predictors. Journal of the American Statistical Association 85:652–663).

**Key Words**: Attenuation; Logistic regression; Measurement error bias; Mixed-effects regression; Regression dilution bias.

## 1. INTRODUCTION

In epidemiologic studies, investigators are often concerned about correcting risk estimates for the bias due to measurement error in the risk factor. While much work has been done that is related to this problem (Carroll and Stefanski, 1990; Carroll et al., 1995; Irwing et al., 1991; Lagakos, 1988; Prentice, 1982; Rosner et al., 1989, 1990; Rosner et al., 1992; Vollmer, 1988), fewer studies have demonstrated how random fluc-
tuations in baseline measurements of risk factors can result in substantial underestimates of the association between the risk factor and the outcome variable (Gardner and Heady, 1973; MacMahon et al., 1990). For example, Rosner and colleagues (1992; 1989; 1990) examined the problem of the underestimation of risk for logistic regression models by developing methods to correct for measurement error. Carroll et al. (1995) discuss measurement errors in nonlinear models and develop a variety of methods for addressing the problem. Carroll and Stefanski (1990) provide a general formulation for measurement error models.

This underestimation of the true association between a risk factor and a disease outcome has been termed regression dilution bias (MacMahon et al., 1990). This bias can arise since an individual’s measurement of a risk factor may differ from its true level due to random variability in the measurement process or due to real but short-term biological variability. Because of regression toward the mean, observed low values are probably lower than the true value of the risk factor, while observed high values are probably higher than the true value.

Figure 1 demonstrates how this regression toward the mean can affect the relationship between a risk factor and the probability of an outcome that is described by a logistic regression model. The apparent shift toward the mean in the levels of the risk factor used in the logistic analysis shows that the “true relationship” between the risk factor and the probability of disease is stronger than is estimated from the observed imperfect data.

One purpose of this paper is to present a statistical approach using a regression calibration/imputation type method for providing risk factor

![Figure 1](image-url)

**Figure 1.** Relationship between the value of a risk factor and the risk of a related disease based on the true values of the risk factor and those predicted from observed values. Regression toward the mean can result in the underestimate of the true strength of the relationship between a risk factor and the probability of disease.
estimates that are closer to the true values for each individual than the actual observed value. This, in turn, reduces the bias in the estimated association between a risk factor measured at baseline and the study’s outcome of interest. An empirical Bayes approach for estimating the true blood pressure values based on mixed-effects regression models (Laird and Ware, 1982; Lindstrom and Bates, 1988) is compared to the observed baseline value or a simple mean of multiple observed values. A second purpose of this paper is to examine the number of repeated observations and/or length of follow-up needed to adequately assess the true baseline level of a risk factor. The resulting recommendation is supported by a simulation study of the different patterns of observations considered in this paper. The new method presented in this paper is compared to the existing measurement error correction method of Rosner et al. (1992). In addition, the standard errors from the new method are compared to a nonparametric bootstrap estimate of the standard errors as well as to the asymptotic standard errors obtained using a method proposed by Carroll and Stefanski (1990).

2. METHODS

The Baltimore Longitudinal Study of Aging (BLSA) follows a group of community-dwelling volunteers who are scheduled to return to the study once every two years for two to three days of tests. Usually data sets from the BLSA represent about 1000 individuals and are characterized by longitudinally-collected and correlated measurements. Various methods have been proposed for analyzing longitudinal data; Diggle et al. (1994) provide an overview of some of these methods. In this article, we use the linear mixed-effects model (LME) presented by Laird and Ware (1982) and Lindstrom and Bates (1988). These models have been applied to a number of BLSA data sets (Morrell and Brant, 1991; Morrell et al., 1995; Pearson et al., 1994).

2.1. Mixed-Effects Model

The LME model for the vector of data for each of the $N$ individuals in the study is

$$y_i = X_i \beta + Z_i b_i + e_i, \quad i = 1, \ldots, N$$

where $y_i$ is the $n_i \times 1$ vector of observations for individual $i$, and $X_i$ and $Z_i$ are the design matrices for the fixed ($\beta$) and random ($b_i$) effects.
Mixed-Effects Models and Measurement Error

The random effects account for the variability among persons and the interaction of persons with the independent variables in the model (Laird et al., 1987). The $b_i$ are estimated as posterior means, and thus the estimates of the $b_i$ are empirical Bayes “shrinkage” estimators.

2.2. Logistic Model

The logistic model is used to estimate the incidence or probability of the disease occurring given the values of s risk factors, $x_{R1}, x_{R2}, \ldots, x_{Rs}$. If the disease outcome is represented by the binary variable

$$Y = \begin{cases} 1, & \text{if disease occurs during the follow-up period} \\ 0, & \text{otherwise}, \end{cases}$$

then the probability of the disease occurring given the risk factors is modeled using the logistic equation

$$P(Y = 1|X_R) = 1/(1 + \exp(-X_R^T \beta_R))$$

where $X_R$ is the $s \times 1$ vector of risk factors and $\beta_R$ is the corresponding vector of parameters. Parameter estimates for the logistic model can be obtained using a maximum likelihood approach (Hosmer and Lemeshow, 1989).

2.3. Determining the Baseline Value of a Risk Factor

Studies that examine the relationship between a set of risk factors and the occurrence of an outcome (disease) rely on baseline measurements that may often be only a single determination of each risk factor. Figure 1 summarizes a potential bias of using a single measurement to assess risk that can occur if that single value is measured with error or has short-term biological error. The implications of blood pressure variability in regard to hypertension lead Rosner and Polk (1979) to consider how many visits or separate examinations and how many measurements per visit should be made in order to get a correct blood pressure classification. Based on repeated visits with multiple measurements per visit over a relatively short period of time, Rosner and Polk (1983) then used the individual’s average measurement to define their true underlying level of blood pressure.

Thus, one way to protect against this regression dilution bias is to use a study where replicated measurements are collected near baseline. These
data could be modeled using the mixed-effects model to obtain a baseline predicted value. These mixed-effects estimates of the baseline risk factor values for each individual are empirical Bayes “shrinkage” estimates which give “much more reasonable” Breslow et al. (1990) estimates over ordinary least squares estimates. In this study, these shrinkage estimates are obtained by using a mixed-effects model that has random-effects terms for the intercept, time (in the study), and time² as variables in the model. Besides the fixed-effects terms corresponding to these random-effects terms (intercept, time, time²), additional fixed-effects are included in the model that are important in estimating individual risk factor levels (namely, terms for age at first examination and obesity level). Finally, the mixed-effects baseline estimates for each individual are imputed as the value of the risk factor in a logistic regression analysis.

For this article, systolic blood pressure measurements that have been collected on 901 BLSA male participants are considered for examining the risk of CHD occurring. Only examinations where each participant was not known to be taking medications that affected blood pressure or where no evidence that CHD was present are used in the analysis. CHD is defined as the occurrence of a coronary death, the diagnosis of a myocardial infarction by history or pathologic Q-waves, or angina pectoris. To estimate the incidence of CHD as a function of a baseline determination of systolic blood pressure, an investigator could use an observed single measurement, or if available, a mean of multiple measurements taken from the first visit examination. There was on average 3 blood pressure measurements (with a maximum of 4) taken for each person in the BLSA sample used in this article. However, if serial data has been collected for at more than one visit with only one measurement per examination, then a baseline value of the risk factor could be represented using a predicted value from some regression model. In particular, one might use the estimated intercept values from a mixed-effects analysis on data collected from at least two repeated visits near baseline. If \( y_{ij} \) represents the \( j \)th blood pressure reading for the \( i \)th study participant, a mixed-effects model such as

\[
y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + (\beta_2 + b_{2i})t_{ij}^2 + \beta_3a_{i1} + \beta_4a_{i2} + \beta_5ob_{i} + \epsilon_{ij}
\]

could be used to estimate each baseline reading. The terms time, age1, and obese represent the person’s time (in years) in the study, age (in years) at first examination, and the obesity status (obese = 1 if body mass index \( \geq 30 \) kg/m², = 0 otherwise), respectively. The terms for time, age1 and obesity were included because they are important predictors of blood
pressure in this data set and because a proper specification of the mixed-effects model is important for obtaining appropriate individual shrinkage estimates. Thus, for comparative purposes, logistic regression analyses of the association between systolic blood pressure and the risk of CHD are calculated using baseline values of blood pressure obtained from: (1) a single observation at the first visit; (2) the mean of multiple observations at the first visit; the linear mixed-effects (LME) estimate of the blood pressure at the first visit based on (3) multiple observations at the first visit; (4) a single observation taken at each of two visits within three years; (5) multiple observations at two visits within three years; (6) single observation taken at each of three visits within five years; and (7) multiple observations at three visits within five years.

Table 1 gives the means and standard deviations of the baseline SBP values for the seven different methods of determining the covariate in the logistic regression model. The table shows that the mean of the SBP baseline values for the 571 non-CHD events are, of course, lower than for the 330 CHD events. In addition, the table shows that the mean values using the LME predictions are quite similar to the corresponding mean when using the observed data for the non-events and events, respectively. Since the most extreme values of the covariates will be shrunk towards the mean by the LME predictions, the standard deviations of the baseline SBP values obtained using the LME models are smaller than for the actual observed data.

The LME model used to describe the data contains longitudinal terms that are appropriate for the number of repeated observations for that data.
set. In particular, when the LME model is used to obtain predictions at first visit based on multiple observations at first visit (3 above), the LME model for describing the data has only one random effect (for intercept) and is

\[ y_{ij} = (\beta_0 + b_{0i}) + \beta_3 \text{age}_i + \beta_4 \text{age}_i^2 + \beta_6 \text{obese}_i + \epsilon_{ij}. \]

If the data set contains observations at two time points (4 and 5 above), the model contains two random effects (for intercept and time) and is

\[ y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \text{time}_{ij} + \beta_3 \text{age}_i + \beta_4 \text{age}_i^2 + \beta_5 \text{age}_i \times \text{time}_{ij} + \beta_6 \text{obese}_i + \epsilon_{ij}. \]

Finally, if the data set contains observations at three time points (6 and 7 above), the LME model for the data includes three random effects (for intercept, time and time\(^2\)) and is given by Eq. (2.1).

Armstrong et al. (1989) presented a method for addressing measurement error in the logistic model for case-control data where there are repeated measurements on the covariate for each individual and the covariate for each individual is subject to error. They assume that the true covariate values among subjects follow separate normal distributions for the cases and controls, where the distributions have different means but a common variance–covariance matrix. The LME model is flexible enough to handle this approach by including event group in the LME model.

To assess the accuracy of the standard errors of the parameters from the logistic regression based upon imputed blood pressure values, a non-parametric bootstrap was performed for each of the LME imputation methods. The standard deviations of the logistic parameter estimates from the 1000 bootstrap samples may be compared to the standard errors produced by the logistic regression based on the imputed values.

### 2.4. Monte Carlo Study

Since different ways of expressing the baseline value of a risk factor \(x_R\) in the logistic model can lead to different estimates of the risk of the outcome, which estimate of the logistic parameter \(\beta_R\) is closer to the truth? To examine this question a Monte Carlo simulation study is used to investigate the bias in logistic parameter estimates. So that the sample data in the simulation study matches the actual data reasonably well, the same design points are used for each individual in the simulation study as are in the actual data as well as the same number of repeated
observations at each point. Thus, the data is generated based on the model for multiple observations at three visits within five years.

In the simulation study we assume that the true blood pressures, $x_R$, follow a normal distribution and the probability of CHD given the true blood pressure is given by the logistic function,

$$P(\text{CHD}|x_R = x_R) = \frac{1}{1 + \exp\left(-\left(\beta_{R0} + \beta_{R1} x_R\right)\right)}.$$ (2.2)

The parameter values for the logistic model and the normal distribution are chosen to closely match the actual data. In particular, the normal mean and standard deviation are 118 and 14.5, respectively, and the logistic parameters are $\beta_{R0} = -4.25$ and $\beta_{R1} = 0.031$. The data generation for the Monte Carlo study proceeds as follows. First, sample 900 true blood pressures from $N(118, 14.5^2)$, and conditional on the observed true blood pressures, compute the probabilities of CHD using the logistic model (2.2). Using these probabilities, generate the CHD event data from Bernoulli distributions. Second, to make sure that each individual’s intercept will match the true blood pressure generated in the first step, the random effect for the intercept for each person for use in the mixed-effects model is computed as the person’s true blood pressure minus the model intercept (which is 118 in this example). Using the covariance matrix of the random effects from the model with multiple observations at three visits within five years, we randomly generate the random effects for each person for time and time$^2$ conditional on the value of the intercept random effect. Third, using these random effects, the expected blood pressure at each of the design points in the data set is computed for each subject based on the fixed-effects parameters for longitudinal follow-up from the model for multiple observations at three visits within five years. The additional fixed effects in the model ($\text{age1}$, $\text{age1}^2$, $\text{age1} \times \text{time}$, and obese) will only affect the intercept for each subject. Since the intercept for each subject has been fixed at the appropriate value to give the correct true blood pressure in step 2 above, these additional fixed effects are not necessary to be included in the simulation study. Finally, fourth, normal random noise is added to these expected blood pressures with error standard deviation from the model with multiple observations at three visits within five years and the same number of repeated values at each design point as in the actual BLSA data set.

First, the logistic model is fit based on the true blood pressures generated at step 1. Next, using the simulated data set, the correct subset of data is selected for fitting a logistic model for systolic blood pressure using baseline values obtained from the seven ways stated previously. As with the analyses of the actual data set in the previous section, the LME model
uses one random effect when there is only baseline data included in the model, two random effects when there are observations at up to two time points, and three random effects when there are up to three time points. This process is repeated 500 times for each of the seven ways selected above for representing the baseline data set. For each of the 500 replications, the parameter estimates and their corresponding standard errors are used to calculate 95% confidence intervals. The coverage proportion of these 95% confidence intervals is computed as the proportion of intervals that include the true parameter values.

3. RESULTS

3.1. BLSA Data

Data from 901 male participants of the BLSA are used in this study. At baseline, the age of the participants ranged from 17 to 96 years. This study investigates the probability of the occurrence of CHD within 30 years of the baseline visit. There were 330 cases of CHD (36.6%) observed by the end of this period. Longitudinal systolic blood pressure values within five years of the baseline measurement are used in this study. There was an average of 2.1 visits per participant with a mean of 3.3 observations per visit.

Table 2 gives the results from the logistic regressions for examining the relationship between systolic blood pressure and the occurrence of CHD for the BLSA data. The table shows the parameter estimate and its standard error (SE) corresponding to the risk factor along with the statistical test of significance ($z$-ratio) for developing CHD using baseline data for systolic blood pressure that is obtained by the seven different methods represented by both observed values and predicted values from a LME regression analysis. The table also presents the bootstrap standard errors. These standard errors are very similar to those produced by the logistic regression based on using the predicted values as covariates. The results for a single observed value at first visit give the smallest parameter estimate (SE) of 0.02748 (0.00475), followed by 0.02962 (0.00503) for the mean of multiple values at first visit. Note that the logistic parameter estimates corresponding to the LME predicted baseline values are all greater than the parameter estimates from the observed data, ranging from 0.03275 (0.00549) to 0.04462 (0.00666).

The results presented in Table 2 are summarized in Fig. 2. If only data from the first visit is used (left panel of Fig. 2), there is a slight increase in the effect of systolic blood pressure on CHD if the mean of
Mixed-Effects Models and Measurement Error

Table 2. Results from logistic regression representing the relationship between systolic blood pressure and CHD for the actual BLSA data.

\[
\begin{array}{l|llll}
\text{Baseline data} & \hat{\beta}_{R1} & \text{SE} & \text{Bootstrap SE} & z\text{-ratio} \\
\hline
\text{Observed} & & & & \\
\text{Single value at first visit} & 0.02748 & 0.00475 & & 5.78 \\
\text{Mean of multiple values at first visit} & 0.02962 & 0.00503 & & 5.89 \\
\text{Predicted}^a & & & & \\
\text{Multiple values at first visit} & 0.03275 & 0.00549 & 0.00541 & 5.96 \\
\text{2 Visits within 3 years} & & & & \\
\text{Single values per visit} & 0.04462 & 0.00666 & 0.00706 & 6.70 \\
\text{Multiple values per visit} & 0.03343 & 0.00552 & 0.00541 & 6.06 \\
\text{3 Visits within 5 years} & & & & \\
\text{Single values per visit} & 0.04326 & 0.00656 & 0.00674 & 6.59 \\
\text{Multiple values per visit} & 0.03389 & 0.00555 & 0.00545 & 6.11 \\
\end{array}
\]

\(^a\text{Predicted value at baseline using appropriate linear mixed-effects model.}\)

Figure 2. Logistic regression curves estimated from systolic blood pressure (SBP) baseline measurements based on observed data or predicted from a linear mixed-effects regression model.
first visit observations are used rather than a single observation at first visit (0.02962 vs. 0.02748; an 8% increase), while there is a much larger increase in effect if the predicted value based on multiple values at first visit is used (0.03275; a 19% increase). The logistic curves based on multiple visits (right panel of Fig. 2) show that similar results are obtained with data from 2 or 3 visits. Surprisingly, the logistic regressions yield higher parameter estimates when based on single values per visit (0.04462 and 0.04326; approximately a 50% increase) than when based on multiple values per visit (approximately 0.033; a 20% increase over the single value at first visit estimate). This could be due to the larger amount of shrinkage that would occur with a smaller amount of data.

3.2. Monte Carlo Results

The results of the 500 Monte Carlo simulations are presented in Tables 3 and 4. Since the CHD event data is randomly generated

<table>
<thead>
<tr>
<th>Baseline data</th>
<th>Mean</th>
<th>SD</th>
<th>MSE</th>
<th>Mean of SEs</th>
<th>Coverage proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases (Range)</td>
<td>325.6</td>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True blood pressure</td>
<td>−4.2466</td>
<td>0.6295</td>
<td>0.3963</td>
<td>0.6176</td>
<td>0.934</td>
</tr>
<tr>
<td>Single value at first visit</td>
<td>−3.5146</td>
<td>0.5628</td>
<td>0.8576</td>
<td>0.5489</td>
<td>0.726</td>
</tr>
<tr>
<td>Mean at first visit</td>
<td>−3.9625</td>
<td>0.6075</td>
<td>0.4517</td>
<td>0.5915</td>
<td>0.910</td>
</tr>
<tr>
<td>Predicted valuesc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple values at first visit 2 visits within 3 years</td>
<td>−4.2352</td>
<td>0.6559</td>
<td>0.4304</td>
<td>0.6381</td>
<td>0.928</td>
</tr>
<tr>
<td>Single values per visit</td>
<td>−4.3659</td>
<td>0.7186</td>
<td>0.5298</td>
<td>0.6909</td>
<td>0.938</td>
</tr>
<tr>
<td>Multiple values per visit</td>
<td>−4.2325</td>
<td>0.6525</td>
<td>0.4260</td>
<td>0.6345</td>
<td>0.932</td>
</tr>
<tr>
<td>Multiple values per visit</td>
<td>−4.2151</td>
<td>0.6813</td>
<td>0.4653</td>
<td>0.6586</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td>−4.2340</td>
<td>0.6514</td>
<td>0.4246</td>
<td>0.6345</td>
<td>0.930</td>
</tr>
</tbody>
</table>

\(^a\)True value of \(\beta_{R0} = −4.25\).

\(^b\)True BP generated from normal distribution and CHD events generated from a Bernoulli distribution conditional on the observed SBP.

\(^c\)Calculated from linear mixed-effects model.
conditional on the observed true SBP values, the number of CHD cases varies among the simulation data sets. The mean number of events is close to the number of events (330) in the actual data set (Table 3). A comparison of the means of the logistic regression parameter estimates with their true values indicates that the shrinkage estimates from the LME model provide the smallest bias, a small mean square error (MSE), and coverage proportions that are close to the nominal level. For example, when the covariate is based on shrinkage estimates from multiple measurements at first visit, the mean of the estimates of $\beta_{R1}$ (Table 4) is 0.03087 (SD, 0.00548) with a MSE of $3.004 \times 10^{-5}$. This is nearly identical to the true value of 0.031 (0.4% under estimate). In addition, the baseline data generated from the LME model for multiple values at first visit has a coverage proportion of 0.926 which is close to 0.944, the coverage proportion for the case with no measurement error. Also, from Table 3, the corresponding coverage proportion for the logistic parameter $\beta_{R0}$ is 0.928, which is also close to 0.934. In terms of the parameter estimate bias and MSE, the shrinkage estimates obtained using linear mixed-effects models based on multiple measurements per visit produce estimates of the logistic regression coefficient (Table 4) that are closer to the true value than the results from the single observed value at

<table>
<thead>
<tr>
<th>Baseline data</th>
<th>Mean</th>
<th>SD</th>
<th>MSE ($\times 10^{-5}$)</th>
<th>Mean of SEs</th>
<th>Coverage proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>True blood pressure</td>
<td>0.03094</td>
<td>0.00524</td>
<td>2.748</td>
<td>0.00513</td>
<td>0.944</td>
</tr>
<tr>
<td>Single value at first visit</td>
<td>0.02479</td>
<td>0.00468</td>
<td>6.052</td>
<td>0.00455</td>
<td>0.702</td>
</tr>
<tr>
<td>Mean first visit</td>
<td>0.02855</td>
<td>0.00507</td>
<td>3.164</td>
<td>0.00491</td>
<td>0.912</td>
</tr>
<tr>
<td>Predicted values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple values at first visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 visits within 3 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single values per visit</td>
<td>0.03199</td>
<td>0.00602</td>
<td>3.727</td>
<td>0.00576</td>
<td>0.946</td>
</tr>
<tr>
<td>Multiple values per visit</td>
<td>0.03084</td>
<td>0.00545</td>
<td>2.975</td>
<td>0.00528</td>
<td>0.928</td>
</tr>
<tr>
<td>3 visits within 5 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single values per visit</td>
<td>0.03071</td>
<td>0.00570</td>
<td>3.258</td>
<td>0.00548</td>
<td>0.928</td>
</tr>
<tr>
<td>Multiple values per visit</td>
<td>0.03085</td>
<td>0.00544</td>
<td>2.966</td>
<td>0.00528</td>
<td>0.930</td>
</tr>
</tbody>
</table>

$^a$True value of $\beta_{R1} = 0.031$.

$^b$True BP generated from normal distribution and CHD events generated from a Bernoulli distribution conditional on the observed SBP.

$^c$Calculated from linear mixed-effects model.
first visit (0.02479, a 20% under estimate) or the mean of observed multiple measurements at first visit (0.02855, a 8% under estimate). The results for single or mean of multiple observed SBP values at a single visit have a greater bias and MSE (though do provide slightly smaller standard deviations of the estimates) than the logistic parameter estimates derived from the linear mixed-effects model’s shrinkage estimates of SPB from multiple measurements at 2 and 3 visits (0.5% under estimates).

The 95% confidence interval coverage proportions for $C_1/C_2$ from the LME model estimates from multiple values per visit at 2 and 3 visits within 3 and 5 years are 0.932 and 0.930, respectively. These are almost identical to 0.934, the coverage proportion derived without measurement error (Table 3).

4. COMPARISON OF LME-CORRECTION PROCEDURE WITH THE PROCEDURE OF ROSNER ET AL.

This section compares the procedure developed by Rosner et al. (1992) with the proposed LME-correction method. Rosner et al. (1992) allow for a main study data set as well as repeated measures from a reproducibility substudy. In their reproducibility substudy there are $r$ available observations per subject. It is assumed that these are measured with error and that the within-person measurement error model is $Z = X + \varepsilon$, where $X$ is the true value of the covariate without measurement error. This is essentially the same assumption as for the present repeated observations at first visit scenario. To compare the proposed LME-correction method to the procedure of Rosner et al. (1992), a copy of the program, MERROR, was obtained to fit their model. The authors assume that there are an equal number of observations per subject in the reproducibility substudy and that this set of data is distinct from the actual covariate. The MERROR program reflects this assumption. The LME-correction method allows for unequal numbers of repeated observations per subject. In the LME model used to estimate the covariate value to be imputed into the logistic regression model, the amount of shrinkage to account for measurement error will be dependent on the number of observations for that subject. In addition, the LME-correction method does not distinguish between which observation is in the main data set and which are included in the reproducibility study.

To compare the two procedures, three analyses were performed using both methods. In all cases only the first visit data is used. The reproducibility study consists of data from the 663 subjects with exactly four
repeated measurements (3 in the reproducibility substudy). The data sets considered are as follows. First, the first SBP at the initial visit for all subjects is used as the main data set. The remaining three observations from the 663 participants with 4 measurements is the reproducibility data set. For the LME-correction analysis, these data are considered as a single data set with 237 subjects with one observation and 663 subjects with four observations. Second, only the 663 participants who had 4 measurements at the first visit are included. The first observation is the main data set, the remaining 3 are the reproducibility data set. For the LME-correction analysis these data are considered as a single data set with 663 subjects with four observations. Third, the 663 participants who had 4 measurements at the first visit are included (as in 2). However, the last observation is the main data set, the remaining 3 are the reproducibility data set.

Table 5 gives the results of these comparisons. The LME-correction method gives identical results for data sets 2 and 3 since this method does not distinguish between the main and reproducibility data sets. As expected, both the LME and MERROR methods of correction give larger estimates of the parameter than the uncorrected logistic parameter. The standard errors are larger for the two correction methods, but are slightly smaller for the LME-correction method than the MERROR method.

A Monte Carlo simulation study is performed to provide comparisons of the parameter estimates and standard errors between the two procedures. Data is generated to reproduce the two types of data sets above, the first with 900 subjects in the main data set and 663 subjects with an additional 3 observations in the reproducibility data set, and the second using only 663 subjects with four observations to make up the main and reproducibility data sets. The simulation proceeds in a similar fashion to that described earlier. The results are given in Table 6.

Table 5. Parameter estimates (SE) comparing the LME-correction procedure with the MERROR-correction procedure.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Actual SBP (uncorrected)</th>
<th>LME correction</th>
<th>MERROR correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.02744 (0.004754)</td>
<td>0.0336 (0.00567)</td>
<td>0.03481 (0.006052)</td>
</tr>
<tr>
<td>2</td>
<td>0.03000 (0.005637)</td>
<td>0.0359 (0.00658)</td>
<td>0.03812 (0.007189)</td>
</tr>
<tr>
<td>3</td>
<td>0.02675 (0.005541)</td>
<td>0.0359 (0.00658)</td>
<td>0.03472 (0.007218)</td>
</tr>
</tbody>
</table>
For the data set with 900 subjects in the main data set and 663 in the reproducibility data set (all subjects), the results show that the LME and MERROR correction procedures provide almost identical results in terms of the mean of the parameter estimates. Both procedures seem to provide unbiased estimates of the true parameter value (with the LME correction having a slightly smaller MSE) and substantial improvements over the uncorrected (single value) logistic. The same result holds for the data set when only 663 subjects are used. However, the mean of the standard errors is smaller for the LME-correction procedure in both cases. The standard errors for these estimates can be compared to each of the corresponding standard errors from the true blood pressure values before the addition of measurement error. The mean of these standard errors is given in the first row of Table 6. The LME-correction standard errors were closer to the standard errors from the true blood pressures than the MERROR-correction standard errors in all but one of the 500 Monte Carlo replications for the first type of data set and all of the samples for the second type of data set. However, the coverage proportions for both the correction methods are almost identical, are a vast improvement over the uncorrected coverage proportions, and are not significantly different from the nominal 0.95 level.

Table 6. Results of 500 Monte Carlo simulations for the logistic regression parameter $\beta_{R1}$ for systolic blood pressure and CHD.$^b$

<table>
<thead>
<tr>
<th>Baseline data</th>
<th>Mean</th>
<th>SD</th>
<th>MSE ($\times 10^{-5}$)</th>
<th>Mean of SEs</th>
<th>Coverage proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>324.5</td>
<td>14.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Range)</td>
<td>(285–370)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True blood pressure</td>
<td>0.03103</td>
<td>0.00532</td>
<td>2.833</td>
<td>0.00514</td>
<td>0.944</td>
</tr>
<tr>
<td>All Subjects ($N = 900$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-value logistic</td>
<td>0.02481</td>
<td>0.00465</td>
<td>5.996</td>
<td>0.00456</td>
<td>0.702</td>
</tr>
<tr>
<td>LME correction</td>
<td>0.03083</td>
<td>0.00548</td>
<td>3.003</td>
<td>0.00535</td>
<td>0.946</td>
</tr>
<tr>
<td>MERROR correction</td>
<td>0.03070</td>
<td>0.00577</td>
<td>3.338</td>
<td>0.00565</td>
<td>0.948</td>
</tr>
<tr>
<td>Subjects with exactly 4 measurements ($N = 663$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-value logistic</td>
<td>0.02477</td>
<td>0.00533</td>
<td>6.723</td>
<td>0.00531</td>
<td>0.788</td>
</tr>
<tr>
<td>LME correction</td>
<td>0.03084</td>
<td>0.00616</td>
<td>3.798</td>
<td>0.00614</td>
<td>0.960</td>
</tr>
<tr>
<td>MERROR correction</td>
<td>0.03065</td>
<td>0.00658</td>
<td>4.338</td>
<td>0.00659</td>
<td>0.950</td>
</tr>
</tbody>
</table>

$^a$True value of $\beta_{R1} = 0.031$.
$^b$True BP generated from normal distribution and CHD events generated from a Bernoulli distribution conditional on the observed SBP.
$^c$Carroll and Stefanski standard errors.
5. DISCUSSION

Measurement error in the explanatory variables in a risk factor study can lead to an underestimation of the true association between the risk factor and the study’s outcome. Rosner et al. (1992; 1989; 1990), Gardner and Heady (1973), and MacMahon et al. (1990) have considered this problem and examined what constitutes an adequate baseline measurement of a risk factor. Gardner and Heady (1973) conclude that the estimate of the true value of an individual’s risk factor is more precise as more observations are made on the individual. They demonstrate that there is little difference between the observed and true values as the within-person variance is reduced. MacMahon and colleagues (1990) show that it is possible to estimate the real association between a person’s usual value of a risk factor and the subsequent occurrence of an event. This is achieved by combining data on the association between the usual value and the baseline value of the risk factor with data on the association between the baseline value of the risk factor and the subsequent occurrence of the event. These authors illustrate that additional measurements of a risk factor that are made several years after a baseline measurement can be used indirectly to correct the risk factor/outcome association. Rosner et al. (1992) procedure for correction of measurement error in logistic regression assumes that two sets of data exists, namely data from the main study population as well as a validation sample or a reproducibility substudy sample which is distinct from the main study sample. The validation or reproducibility samples are used to obtain factors that can be used to make corrections to the logistic parameters to account for the measurement error.

Rosner et al. (1992) correct for attenuation by applying a correction to the naive estimator (the logistic parameter estimate based on the observed covariate data). The method presented in this paper follows the regression calibration ideas of Carroll et al. (1995) by using the available replicate data to estimate the expected value of the true covariate (without error) for each participant conditional on the observed covariates (with error). In the simple case of using replicate data, their best linear approximant of the true covariate is given as a weighted average of the overall mean of the observed covariates and the individual participant’s mean. The weights are based on within- and between-subject sums of squares. For participants with large numbers of repeated observations, more weight is given to that participant’s mean, while for participants with few observations more weight is given to the overall mean. This is very similar to the present method of repeated values at first visit where the estimate based on the LME model is also given as a weighted average of the overall estimate and the individual’s mean.
This paper considers several ways of obtaining the baseline value of the explanatory variable in a risk factor study. The number of observations needed and the method of correcting for the measurement error are studied. Besides using observed data from single and multiple measurements collected at first visit, predicted baseline values from a mixed-effects model using data within a few years of baseline are used to estimate each individual's risk factor value. Actual or predicted systolic blood pressure is used as a predictor in the study of the development of CHD using a logistic regression model for data collected from the BLSA. The estimated values of the logistic parameter, \( \beta_{R1} \), and their corresponding tests of significance are smallest using both a single observed value and the mean of observed multiple values at first visit, while the estimated coefficients and their corresponding significance are greater using predicted baseline blood pressure values from a linear mixed-effects model (Table 2). Surprisingly, single measurements from multiple visits yield higher logistic regression parameter estimates than multiple measurements at multiple visits. This may be due to the larger amount of shrinkage that would occur with a smaller amount of data. However, in this data set we do not know the "true" association so we cannot assume that larger coefficients are necessarily better.

A simulation study was carried out to examine which method of representing the baseline data gives the most accurate results. The Monte Carlo simulation study showed that shrinkage estimates from the linear mixed-effects model based on multiple measurements per visit yielded the more accurate and robust estimates of the true risk relationship between systolic blood pressure and CHD than observed data or shrinkage estimates from single measurements per visit (Table 4). A single observed SBP value underestimates the true logistic regression coefficient by 20%. Taking the simple arithmetic mean of multiple observed SBP values reduced the underestimation to 8%. However, the LME shrinkage estimates obtained from these same observations yielded an estimate of the logistic regression coefficient that was within 1% of the true value and was one of the most robust approaches as evidenced by one of the lowest mean square errors. In general, the mean square error for \( \beta_{R1} \) is smallest for the shrinkage estimates from the mixed-effects model based on multiple measurements per visit. Note that the mean square error for models based on single measurements per visit are larger than their corresponding mean square error values based on multiple visits, presumably because analyses based on multiple measurements per visit are less sensitive to influential outliers. As with the actual estimates in Table 2, the single values at first visit provide the largest estimates of the logistic parameter. Thus, if multiple observations are available they should be used to obtain the best
estimates of the effect of the risk factors. This is merely an extension of the common practice of using the mean of 2 or 3 blood pressure measurements in many epidemiological and clinical studies.

Our approach for correcting for bias due to measurement error can be viewed as an imputation method. A drawback of these methods is that the variance of the corrected estimate fails to take into account the uncertainty due to the estimation of the covariate. However, our bootstrap standard errors and the Monte Carlo simulation study show that the standard errors readily obtained from the LME-correction method accurately reflect the uncertainty in the logistic parameter estimates, provide inferences with the desired coverage proportions, and are close to the standard errors for the data without measurement error. Another possible approach to this problem is to use the multiple imputation method developed by Rubin (1987). This method would take into account the uncertainty due to the estimation of the imputed value and it is perhaps a worthwhile avenue for future study.

Alternatively, Carroll and Stefanski (1990) develop M-estimators for a general formulation of measurement error models. Using their expressions, standard errors for the regression coefficients in the logistic model are derived for the repeated values at first visit case (see Appendix) and are compared to the simulated values in Table 6. These standard errors are almost identical to those computed from the LME-correction method.

The mixed-effects model approach described in this paper for correcting for measurement error requires that the marginal distribution of the response variable be normal. The simulation also relies on this assumption as the data is randomly generated from a normal distribution. In the example presented, this assumption appears plausible. However, this normality assumption may not be true in general. One approach is to find a transformation of the response variable so that the normality assumption on the error components will be satisfied and apply our method using the transformed data. One could also investigate the effect of the normality assumption by conducting further simulation studies in which the data are generated from a variety of distributions. The properties of the estimators and their inferences could then be considered. However, this is beyond the scope of this current paper and may be investigated at a later stage.

6. CONCLUSION

This paper has discussed the selection of appropriate baseline values of a risk factor in the assessment of the effect of the risk factor on an
outcome variable using logistic regression. If a study has already been performed and only a single value of the risk factor is available at baseline then estimates of the effect of the risk factor on the outcome variable will be underestimated due to regression dilution bias.

If multiple values of the risk factor are available at baseline then improved estimates of the logistic parameter can be obtained by using shrinkage estimates of the baseline values of the covariate from a linear mixed-effects model with a fixed and random effect for intercept. In the case when single values are available at a number of follow-up times, improved estimates (over a single baseline value) of the logistic parameter are obtained by using an appropriate LME model to predict the baseline covariate. If multiple values are available at a number of points, relatively little is gained by using all the data compared to an LME model based on first visit data.

Finally, if a study is to be designed, the simulation would suggest that having multiple measurements at baseline and using the LME model to predict the baseline covariate values for the logistic model provides for logistic parameter estimates with small MSE. Such an approach using multiple baseline measurements is less complex and relies on fewer additional assumptions than do using more complex LME models with additional design points. This LME approach to adjust for measurement error is easily implemented in statistical packages (e.g., BMDP, SAS, SPLUS), is flexible in dealing with data sets that are unbalanced in the number and spacing of repeated measurements, and can be easily generalized to other types of clustered data (e.g., households, families, regions, etc.).

APPENDIX

In this appendix we apply the methods of Carroll and Stefanski (1990) to obtain standard errors of our LME-corrected logistic parameter estimates for the case of repeated values at first visit. Following the notation of Carroll and Stefanski (1990), let $Y$ be the outcome variable, $X$ the true value of the explanatory variable without error, and $W$ the value of the explanatory variable measured with error. Then $Y|X \sim\text{Bin}(1, p)$ where the conditional mean is $p = E(Y|X = x) = f_m(x, \beta) = 1/(1 + e^{-(\beta_0 + \beta_1 x)})$ and $\text{Var}(Y|X = x) = f_v(x, \beta) = e^{-(\beta_0 + \beta_1 x)}/(1 + e^{-(\beta_0 + \beta_1 x)})^2$. We assume that the measurement error model is given as $W = X + \delta U$ where $\text{Cov}(U) = I$. Consequently, the repeated values for a particular $X$ will be independent with measurement error variance given by $\delta^2$. Our data consists of $n$ observations $(Y_i, W_{i1}, \ldots, W_{ik})$. Thus, we have $n$ internal reliability observations.
with possibly varying numbers of repeated measurements per subject. Our strategy to account for measurement error is to replace $X$ by its estimate obtained from the LME model, $\hat{x}_i$. Then

$$F_i = \begin{pmatrix} y_i - (1/(1 + e^{-(\beta_0 + \beta_1 \hat{x}_i)})) \\ \hat{x}_i (y_i - (1/(1 + e^{-(\beta_0 + \beta_1 \hat{x}_i)}))) \\ -k_i/(2\delta^2) + \left( \sum_{j=1}^{k_i} (w_{ij} - \hat{x}_i)^2 \right)/(2(\delta^2)^2) \end{pmatrix}.$$  

The first row corresponds to the parameter $\beta_0$, the second row corresponds to the parameter $\beta_1$, and the last row corresponds to the parameter $\delta^2$. Then $\hat{B} = \frac{1}{n} \sum_{i=1}^{n} \bar{F}_i \bar{F}_i^T$ where the $\bar{F}_i$ are obtained by inserting the logistic parameter estimates for the $\beta$'s and the variance term from the LME model as an estimate of $\delta^2$ into $F_i$. Next

$$G_i = \begin{pmatrix} -e^{-(\beta_0 + \beta_1 \hat{x}_i)} & -\hat{x}_i e^{-(\beta_0 + \beta_1 \hat{x}_i)} & 0 \\ (1 + e^{-(\beta_0 + \beta_1 \hat{x}_i)})^2 & (1 + e^{-(\beta_0 + \beta_1 \hat{x}_i)})^2 & 0 \\ -\hat{x}_i^2 e^{-(\beta_0 + \beta_1 \hat{x}_i)} & -\hat{x}_i^2 e^{-(\beta_0 + \beta_1 \hat{x}_i)} & 0 \end{pmatrix}$$

and $\hat{A} = \frac{1}{n} \sum_{i=1}^{n} \bar{G}_i$, where the $\bar{G}_i$ are obtained by inserting estimates of the parameters into $G_i$, as before. Finally the asymptotic covariance matrix can be estimated by

$$\text{Cov}(\beta_0, \beta_1, \delta^2) = \frac{\hat{A}^{-1} \hat{B} (\hat{A}^T)^{-1}}{n}.$$  

REFERENCES


Mixed-Effects Models and Measurement Error


