Published in final edited form as:

Biom J. 2016 July ; 58(4): 944–961. doi:10.1002/bimj.201500195.

Estimating covariate-adjusted measures of diagnostic accuracy based on pooled biomarker assessments

Christopher S. McMahan^{*,1}, Alexander C. McLain², Colin M. Gallagher¹, and Enrique F. Schisterman³

¹Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA

²Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

³Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA

Abstract

There is a need for epidemiological and medical researchers to identify new biomarkers (biological markers) that are useful in determining exposure levels and/or for the purposes of disease detection. Often this process is stunted by high testing costs associated with evaluating new biomarkers. Traditionally, biomarker assessments are individually tested within a target population. Pooling has been proposed to help alleviate the testing costs, where pools are formed by combining several individual specimens. Methods for using pooled biomarker assessments to estimate discriminatory ability have been developed. However, all these procedures have failed to acknowledge confounding factors. In this paper, we propose a regression methodology based on pooled biomarker measurements that allow the assessment of the discriminatory ability of a biomarker of interest. In particular, we develop covariate-adjusted estimators of the receiver-operating characteristic curve, the area under the curve, and Youden's index. We establish the asymptotic properties of these estimators and develop inferential techniques that allow one to assess whether a biomarker is a good discriminator between cases and controls, while controlling for confounders. The finite sample performance of the proposed methodology is illustrated through simulation. We apply our methods to analyze myocardial infarction (MI) data, with the goal of determining whether the pro-inflammatory cytokine interleukin-6 is a good predictor of MI after controlling for the subjects' cholesterol levels.

Keywords

AUC; Biological markers; Biomarker pooling; ROC curve; Youden's index

*Corresponding author: mcmaha2@clemson.edu, Phone: +1-864-656-4566; Fax: +1-864-656-5230.

Additional supporting information including source code to reproduce the results may be found in the online version of this article at the publisher's web-site

Conflict of interest

The authors have declared no conflict of interest.

1 Introduction

The statistical evaluation of biological markers (or biomarkers) for the purpose of assessing exposures and disease detection is rapidly increasing in epidemiological research. For example, approximately 70% of the original articles appearing in *Clinical Chemistry* in 2011 focused on biomarker evaluation (Boyd et al., 2012). Epidemiological and medical research journals are replete with studies that evaluate biomarkers' prognostic abilities for various diseases, including HIV (Lau et al., 2009; Kanekar, 2010), cancer (Dhanasekaran et al., 2001; Borges et al., 2013), and cardiovascular disease (Rasmussen et al., 2011; Sabatine et al., 2012). Research in these areas can be limited by the cost associated with assessing new biomarkers; for example, a single analytical measurement of 61 polychlorinated and 13 polybrominated compounds costs \$1400 per specimen (Caudill, 2012). One possible cost saving device is to make biomarker assessments on pooled samples; that is, to measure biomarker levels on pools formed by amalgamating several individual specimens.

Pooled (or group) testing was first proposed by Dorfman (1943) as a means to cost-efficiently screen World War II inductees for syphilis. Recently, group testing methods have branched out into the area of biomarker evaluation using pooled assessments. The key feature of pooling that facilitates a reduction in testing cost is that instead of measuring biomarker specimens one by one, pools (of multiple specimens) are constructed and measured. Biomarker assessments based on pooled samples result in a continuous value, which is typically assumed to represent the average (or sum) of the biomarker concentrations over the pooled individuals (e.g., see Faraggi et al., 2003; Liu and Schisterman, 2003; Schisterman et al., 2004; Mumford et al., 2006; Bondell et al., 2007; Vexler et al., 2008; Malinovsky et al., 2012). The use of pooling can offer a drastic reduction in data collection costs. For example, the 2005–2006 National Health and Nutrition survey reduced the number of analytical measurements required to characterize the distribution of polychlorinated and polybrominated compounds within the population of the United States from 2201 to 228 by pooling, resulting in a savings of \$2.78 million (Caudill, 2012). Further, measured pooled concentrations tend to be larger than the smaller individual concentration levels, thereby reducing the number of measurements subjected to the lower limit of detection (Mumford et al., 2006).

Statistical research in biomarker pooling has predominantly focused on developing methods that can estimate the unadjusted receiver-operating characteristic (ROC) curve and corresponding area under the curve (AUC) through the analysis of pooled biomarker assessments (Faraggi et al., 2003; Liu and Schisterman, 2003; Mumford et al., 2006; Bondell et al., 2007; Vexler et al., 2008; Malinovsky et al., 2012). The ROC curve and AUC are fundamental measures that summarize the discriminatory ability of a biomarker of interest, with respect to correctly distinguishing between cases (diseased) and controls (nondiseased). Along with AUC, Youden's index (Youden, 1950) provides a global measure of performance of a diagnostic test. Youden's index is a measure of the diagnostic effectiveness of a biomarker when the cutpoint (i.e., the value of the biomarker that differentiates cases and controls) is chosen to optimize diagnostic criteria. While the estimation of these summary measures with pooled data has been studied in the literature

(see the references given above), the previous estimation techniques do not allow for covariate adjustment.

It is common in epidemiological analyses to have one or more confounding factors (e.g., age, sex, race, etc.) that are related to both level of a biomarker and prevalence of the condition. In the presence of confounding, unadjusted measures of diagnostic accuracy can reflect the association between the confounders and disease, and not the true diagnostic ability of the biomarker (Janes and Pepe, 2008). Further, a biomarker may possess a great deal of discriminatory power within one stratum of the population, while being completely ineffective in another. Differential discriminatory power can be identified and accounted for through the development of covariate-adjusted measures of diagnostic accuracy and can be utilized to design individualized diagnostic tests (Janes and Pepe, 2009; Huang et al., 2011). It is desirable to adjust for covariate information when estimating ROC curves, AUC, and Youden indices. Until now, this has not been addressed within the pooling literature.

In this paper, we develop methods of estimating covariate-dependent ROC curves, AUCs, and Youden indices based on pooled biomarker assessments. Our estimation technique can be used for modeling biomarker data under various distributional assumptions. Under a Gaussian assumption, we show that the resulting estimates have asymptotic normal distributions and derive asymptotic variance estimates, confidence intervals, and confidence bands. Our methods can be utilized to accurately assess the diagnostic accuracy, and test if a biomarker can significantly discriminate between cases and controls, while controlling for confounding factors. The development of these techniques is motivated by a case-control study conducted at the Cedars-Sinai Medical Center, which consisted of 80 volunteers, half of which had recently experienced a myocardial infarction (MI). For each participant in the study, the pro-inflammatory cytokine interleukin-6 (IL-6) was measured individually, as well as on pools consisting of two and four specimens. In our analyses, we quantify the diagnostic accuracy that IL-6 has in predicting MI status while controlling for cholesterol level, a potential confounder.

The remainder of this paper is organized as follows. In Section 2, we present the general modeling framework for pooled biomarker assessments and demonstrate how our methods can be used for estimating covariate-adjusted ROC curves, AUCs, and Youden indices. In Section 3, we illustrate the finite sample performance of our proposed methodology using simulation. As a part of this study, we introduce two common pooling strategies and investigate their influence on both estimation and inference. In Section 4, we analyze the motivating data and identify cholesterol levels for which IL-6 is a good discriminator of MI status. Section 5 presents concluding remarks and a discussion of future research. Theoretical proofs and derivations are provided in Supporting Information Appendices A and B, along with several additional simulation studies that assess the characteristics of our proposed methodology under various scenarios to include model misspecification.

2 Pooled data modeling

In this section, we introduce a general modeling framework for pooled biomarker assessments, develop methods of evaluating modeling assumptions, and provide techniques for assessing the discriminatory ability of the biomarker under investigation.

2.1 Modeling

For the control (healthy) individuals, let Y_{ij-} , for $i = 1, \dots, c_-$ and $j = 1, \dots, J_-$ denote the biomarker level of the i -th specimen that is assigned to the j -th pool, where c_- and J_- represent the number of individual specimens combined to form a pool and the number of pools, respectively. For modeling purposes, we assume that healthy (nondiseased) individuals' biomarker levels are related to a covariate vector through the following model:

$$Y_{ij-} = \mathbf{X}'_{ij-}\boldsymbol{\beta}_- + \epsilon_{ij-}, \text{ for } i = 1, \dots, c_- \text{ and } j = 1, \dots, J_-, \quad (1)$$

where $\mathbf{X}_{ij-}$ is a p -dimensional vector of covariates and $\boldsymbol{\beta}_-$ is the corresponding vector of regression parameters. We assume that the $\epsilon_{ij-} \stackrel{iid}{\sim} F_-$, where F_- is the cumulative distribution function (CDF) of a real-valued random variable having mean 0 and variance σ_-^2 . Similarly, let, for the case (diseased) individuals, Y_{ij+} denote the biomarker level of the i -th specimen that is assigned to the j -th pool, and we assume that Y_{ij+} is related to a covariate vector through the following model:

$$Y_{ij+} = \mathbf{X}'_{ij+}\boldsymbol{\beta}_+ + \epsilon_{ij+}, \text{ for } i = 1, \dots, c_+ \text{ and } j = 1, \dots, J_+, \quad (2)$$

where $\mathbf{X}_{ij+}$ is a p -dimensional vector of covariates and $\boldsymbol{\beta}_+$ is the corresponding vector of regression parameters. We assume that the $\epsilon_{ij+} \stackrel{iid}{\sim} F_+$, where F_+ is the CDF of a real-valued random variable having mean 0 and variance σ_+^2 . Further, we assume that the ϵ_{ij-} s and ϵ_{ij+} s are independent.

When assessments are made on pools, the individual biomarker levels (i.e., Y_{ij-} and Y_{ij+}) are unobserved (latent) variables. Instead, the observed data consist of an aggregated measurement based on the individual biomarker levels within each pool. Throughout, as is common in the literature, we assume that this aggregated response represents the arithmetic average of the individual biomarker levels. In practice, we view this assumption to be reasonable as long as pools are formed from equal volume aliquots. Under this assumption, the observed pooled assessments are of the form

$$Y_{j-} = \frac{1}{c_-} \sum_{i=1}^{c_-} Y_{ij-} = \bar{\mathbf{X}}'_j \boldsymbol{\beta}_- + \epsilon_{j-}, \text{ for } j = 1, \dots, J_-,$$

$$Y_{j+} = \frac{1}{c_+} \sum_{i=1}^{c_+} Y_{ij+} = \bar{X}'_{j+} \beta_+ + \epsilon_{j+}, \quad \text{for } j = 1, \dots, J_+,$$

for the nondiseased and diseased classes, respectively, where $\bar{X}_{j-} = c_-^{-1} \sum_{i=1}^{c_-} \mathbf{X}_{ij-}$ and $\bar{X}_{j+} = c_+^{-1} \sum_{i=1}^{c_+} \mathbf{X}_{ij+}$. Consequently, ϵ_{j-} and ϵ_{j+} are independent random variables each having mean 0 and variances $\tau_-^2 = c_-^{-1} \sigma_-^2$ and $\tau_+^2 = c_+^{-1} \sigma_+^2$, respectively.

Using the observed data $\{(Y_{j-}, \bar{X}_{j-})\}_{j=1}^{J_-}$ and $\{(Y_{j+}, \bar{X}_{j+})\}_{j=1}^{J_+}$, one can obtain unbiased estimates of the regression coefficients via least squares,

$$\hat{\beta}_- = (\bar{X}'_- \bar{X}_-)^{-1} \bar{X}'_- Y_- \quad \text{and} \quad \hat{\beta}_+ = (\bar{X}'_+ \bar{X}_+)^{-1} \bar{X}'_+ Y_+,$$

where $\bar{X}_- = (\bar{X}_{1-}, \dots, \bar{X}_{J_- -})'$ and $\bar{X}_+ = (\bar{X}_{1+}, \dots, \bar{X}_{J_+ +})'$ are the usual design matrices and $Y_- = (Y_{1-}, \dots, Y_{J_- -})'$ and $Y_+ = (Y_{1+}, \dots, Y_{J_+ +})'$ are the corresponding response vectors. We also obtain unbiased estimates of the error variances,

$$\hat{\sigma}_-^2 = c_- (J_- - p)^{-1} Y'_- (I - H_-) Y_- \quad \text{and} \quad \hat{\sigma}_+^2 = c_+ (J_+ - p)^{-1} Y'_+ (I - H_+) Y_+,$$

where $H_- = \bar{X}_- (\bar{X}'_- \bar{X}_-)^{-1} \bar{X}'_-$, $H_+ = \bar{X}_+ (\bar{X}'_+ \bar{X}_+)^{-1} \bar{X}'_+$, and I is the matrix. Further, given the the resulting regression model, standard statistical software can be used for obtaining these estimators, as well as to conduct typical regression diagnostics, model building, hypothesis testing, and inference.

2.2 Biomarker assessment

A key aspect in the epidemiologic study of biomarkers is to assess their discriminatory ability in differentiating between diseased and nondiseased subjects. Discriminatory ability is commonly measured as the predictive accuracy of a test that diagnoses a person as positive (negative) if the level of their measured biomarker is above (below) a threshold. The diagnostic accuracy of a test with threshold t , is typically quantified through its sensitivity and specificity, denoted by $S_e(t)$ and $S_p(t)$, respectively. The sensitivity (specificity) of a test is defined as the probability that the test will correctly diagnose an individual as positive (negative). Without loss of generality, we assume throughout that a positive diagnosis is rendered if a patient's biomarker level exceeds a predetermined threshold, and negative otherwise.

In the presence of confounding factors, covariate-adjusted values of sensitivity and specificity are defined as $\text{pr}(Y_+ > t|X)$ and $\text{pr}(Y_- < t|X)$, respectively, where $Y_+(Y_-)$ represents the biomarker level for a randomly selected positive (negative) individual having

covariate X . Under the modeling assumptions in Section 2.1, covariate-adjusted sensitivities and specificities can be expressed as

$$S_e(X, t, \theta) = 1 - F_+(t - X'\beta_+) \text{ and } S_p(X, t, \theta) = F_-(t - X'\beta_-),$$

respectively, where $\theta = (\beta'_+, \beta'_-, \sigma_+^2, \sigma_-^2)'$. Thus, conditional ROC curves given X can be constructed by plotting $S_e(X, t, \theta)$ versus $1 - S_p(X, t, \theta)$ across all feasible values of t . These curves can be utilized to assess discriminatory ability at particular predictor levels. However, they do not provide a succinct global assessment of the discriminatory ability of a biomarker across the entire predictor space.

In what follows, we present two covariate-dependent measures that provide a more comprehensive summary of the discriminatory ability of a biomarker of interest. The aforementioned covariate-dependent sensitivity and specificity vary not only with X , but also with t . In practice, the threshold value of a diagnostic test would be selected to optimize its predictive ability. As a result, a single summary measure of the discriminatory ability of a biomarker is given by Youden's index, which at a specific predictor level can be defined as

$$YI(X, \theta) = \sup_{t \in \mathbb{R}} \{S_e(X, t, \theta) + S_p(X, t, \theta) - 1\}.$$

Youden's index takes on values between 0 and 1, with values close to 1(0) indicating that the biomarker is effective (ineffective) at discriminating between cases and controls. At a particular predictor level X , the value of t , for example $t_0 = t_0(X, \theta)$, that achieves the supremum is referred to as the optimal threshold (or cutpoint). Under certain distributional assumptions, a closed-form expression for t_0 is available (Schisterman et al., 2005). Further, once the optimal cutpoint is determined, Youden's index can be computed as $YI(X, \theta) = \{S_e(X, t_0, \theta) + S_p(X, t_0, \theta) - 1\}$. The conditional Youden index is a summary measure of the information that is provided in the conditional ROC curve, which simultaneously identifies the optimal cutpoint.

A second measure that summarizes the information in an ROC curve is the conditional AUC, denoted by $AUC(X, \theta)$. In particular, $AUC(X, \theta)$ represents the conditional probability, given X , that a randomly chosen case's biomarker level will be greater than a randomly chosen control, that is, $\text{pr}(Y_+ > Y_- | X)$. Under the modeling assumptions in Section 2.1, one can express $AUC(X, \theta)$ as

$$AUC(X, \theta) = 1 - \int_{\mathbb{R}} F_+\{X'(\beta_- - \beta_+) + \epsilon\} f_-(\epsilon) d\epsilon,$$

where f_- is the probability density function corresponding to F_- . Like Youden's index, $AUC(X, \theta)$ provides a concise evaluation of the discriminatory ability of a biomarker at a given predictor level. $AUC(X, \theta)$ takes on values between 0.5 and 1, with values close to 1(0.5) indicating that the biomarker is effective (ineffective) at distinguishing between the cases and controls.

2.3 Estimation, asymptotic properties, and inferential techniques

Motivated by our data application (see Section 4 and Faraggi et al. (2003) for related details), we proceed under the assumption that the error terms (i.e., ϵ_{ij-} and ϵ_{ij+}) follow a normal distribution. This assumption is common in the pooled biomarker assessment literature (Faraggi et al., 2003; Liu and Schisterman, 2003; Liu et al., 2004; Mumford et al., 2006). Further, under other distributional assumptions, similar estimators and their asymptotic properties could be established in a similar fashion as the approach described herein. Under this assumption, one can express the covariate-adjusted sensitivities and specificities as

$$S_e(X, t, \theta) = \Phi\left(\frac{X'\beta_+ - t}{\sigma_+}\right) \quad \text{and} \quad S_p(X, t, \theta) = \Phi\left(\frac{t - X'\beta_-}{\sigma_-}\right),$$

where Φ denotes the CDF of the standard normal distribution. Using these expressions, it is possible to obtain a closed-form expression for the optimal cutpoint, and hence Youden's index at a particular predictor level X and for a given value of θ (for further details, see Supporting Information Appendix A). Similarly, the conditional AUC is given by

$$\text{AUC}(X, \theta) = \Phi\left(\frac{X'\beta_+ - X'\beta_-}{\sqrt{\sigma_+^2 + \sigma_-^2}}\right).$$

Estimates of the covariate-dependent sensitivities, specificities, ROC curves, optimal cutpoints, Youden indices, and AUCs can be constructed by replacing θ with $\hat{\theta} = (\hat{\beta}'_+, \hat{\beta}'_-, \hat{\sigma}_+^2, \hat{\sigma}_-^2)'$ in the above mentioned expressions. The covariate-dependent optimal cutpoint, Youden's index, and AUC provide a more comprehensive summary of the biomarker of interest across the predictor space. In Supporting Information Appendix B, we establish the asymptotic normality of these estimators, that is, we show that at a given predictor level X , as $J \rightarrow \infty$,

$$\sqrt{J}\{\widehat{\text{YI}}(X, \hat{\theta}) - \text{YI}(X, \theta)\} \xrightarrow{d} N(0, \Sigma_{\text{YI}}),$$

$$\sqrt{J}\{\widehat{t}_0(X, \hat{\theta}) - t_0(X, \theta)\} \xrightarrow{d} N(0, \Sigma_{t_0}),$$

$$\sqrt{J}\{\widehat{\text{AUC}}(X, \hat{\theta}) - \text{AUC}(X, \theta)\} \xrightarrow{d} N(0, \Sigma_{\text{AUC}}),$$

where it is assumed, for the ease of exposition, that $J_- = J_+ = J$. Closed-form expressions for the asymptotic variances (i.e., Σ_{YI} , Σ_{t_0} , and Σ_{AUC}) are provided in Supporting Information

Appendix B, along with their corresponding finite sample estimators, which allow for the situation in which $J_- = J_+$.

To provide a more thorough analysis at a particular covariate level X , one could construct $(1 - \alpha)100\%$ pointwise confidence intervals for $t_0(X, \theta)$, $YI(X, \theta)$, and $AUC(X, \theta)$ using the above mentioned asymptotic results. In Supporting Information Appendix B, we provide expressions for these confidence intervals, along with their derivation. It is important to note that these pointwise confidence intervals are valid for a single fixed value of X at which inference is being drawn. While making multiple simultaneous comparisons (i.e., drawing inference at several levels of X), appropriate corrections to the significance level should be made in order to control the family-wise error rate (Hochberg and Tamhane, 1987). To circumvent this issue, the discriminatory ability of a biomarker can also be simultaneously assessed across the entire covariate space by constructing asymptotic $(1 - \alpha)100\%$ confidence bands for $AUC(X, \theta)$, that is, the sets $C(X)$ can be constructed such that

$$\text{pr}\{AUC(X, \theta) \in C(X) \text{ for all } X\} = 1 - \alpha.$$

In Supporting Information Appendix B, we show that the sets of this form can be constructed as

$$C(X) = \left[\Phi \left(\frac{X'(\hat{\beta}^+ - \hat{\beta}^-)}{\sqrt{\hat{\sigma}_+^2 + \hat{\sigma}_-^2}} - \sqrt{\chi_{p, \alpha}^2} \sqrt{\hat{\Sigma}_{AUC*}} \right), \Phi \left(\frac{X'(\hat{\beta}^+ - \hat{\beta}^-)}{\sqrt{\hat{\sigma}_+^2 + \hat{\sigma}_-^2}} + \sqrt{\chi_{p, \alpha}^2} \sqrt{\hat{\Sigma}_{AUC*}} \right) \right],$$

where $\chi_{p, \alpha}^2$ denotes the $(1 - \alpha)$ -th quantile of a chi-squared distribution with p degrees of freedom, and $\hat{\Sigma}_{AUC*}$ is an asymptotic variance estimator whose explicit form is provided in Supporting Information Appendix B.

The above confidence bands have practical relevance for epidemiological studies interested in testing if a biomarker can significantly discriminate between cases and controls. For example, if one is interested in testing if a biomarker is a significant predictor of disease for all (any) X , then one could check whether or not $C(X) > 0.5$ for all (any) X . Similarly, tests can be conducted to determine whether a new biomarker is a better predictor of disease than a historical biomarker with known AUC. Further, analogous assessments for a given subgroup could be performed using the pointwise confidence intervals.

3 Simulation study

To illustrate the finite sample performance of our proposed methodology, we have designed the following simulation study. Individual biomarker levels were simulated according to the following two models:

$$\text{Case model: } Y_{k+} = \mathbf{X}'_{k+} \boldsymbol{\beta}_+ + \epsilon_{k+} \quad \text{for } k = 1, \dots, N,$$

$$\text{Control model: } Y_{k-} = \mathbf{X}'_{k-} \boldsymbol{\beta}_{-} + \epsilon_{k-} \quad \text{for } k = 1, \dots, N,$$

respectively, where $\boldsymbol{\beta}_{+} = (1.750, 0.015)'$ and $\boldsymbol{\beta}_{-} = (3.000, -0.005)'$, $\mathbf{X}_{k+} = (1, X_{k1+})'$, $(1, X_{k1-})'$. The covariates and error terms were independently sampled as follows: $X_{k1+} \sim \mathcal{N}(225, 40^2)$, $X_{k1-} \sim \mathcal{N}(205, 40^2)$, $\epsilon_{k+} \sim \mathcal{N}(0, 2.15^2)$ and $\epsilon_{k-} \sim \mathcal{N}(0, 1.35^2)$. These models were chosen to emulate the structure of the data analyzed in Section 4. In Supporting Information Appendix C, we provide three additional simulation studies, to include two robustness studies, which further examine the performance of our proposed methodology under a variety of other scenarios.

In order to create pooled data, we chose to investigate two pooling schemes that are prominent among the group testing literature, that is, random and homogeneous pooling, denoted by RP and HP, respectively. Briefly, RP assigns individuals to pools at random, without regard to their covariate values, while HP performs pool assignment so that individuals belonging to a common pool have similar covariate values (for further discussion, see Vansteelandt et al., 2000; Bilder Tebbs, 2009). In either case, once the pool assignment is complete, a pool's measured biomarker level is determined as the mean of the biomarker levels of the individuals assigned to it. We specified that $c_{+} = c_{-} = c$, where $c \in \{1, 2, 4\}$ corresponding to individual testing, pools of size 2, and pools of size 4, with $N \in \{40, 80, 160\}$. Note that the biomarker levels of the N individuals are assessed via $J = N/c$ tests. As a result, using $J = 40$ tests corresponds to $c = 1$ when $N = 40$, $c = 2$ when $N = 80$, and $c = 4$ when $N = 160$. That is, pooled assessments based on $N = 80(160)$ individuals can be made using pools of size $c = 2(4)$ at approximately the same cost of making subject level assessments on $N = 40$ individuals. For each configuration of (c, N) and pooling scheme, we generated 10,000 independent data sets under both case and control models. We then analyzed the data using the methodology outlined in Section 2.

Tables 1 and 2 summarize the parameter estimates based on HP and RP, respectively, for both case and control models, across all combinations of (c, N) . These results indicate that our approach can accurately estimate all parameters across all considered configurations. Estimation accuracy and efficiency appear to be universally better under the HP strategy, when compared to the RP strategy. This finding is congruous with the results presented in the binary group testing regression literature (cf. Zhang et al., 2013; Wang et al., 2014). Therefore, we will primarily focus the discussion herein on the results pertaining to the HP strategy.

A primary objective of this study involves comparing the estimation efficiency and inferential characteristics obtained through analyzing pooled and individual data. For a given value of N , under HP, one will note that the estimated regression coefficients resulting from individual data and pooled data accurately estimate the true underlying regression parameter, but more interestingly the loss in efficiency for these parameter estimates due to pooling is negligible. In contrast, a minor loss of efficiency is realized when trying to estimate the error variances (i.e., σ_{+}^2 and σ_{-}^2) from pooled data, when compared to estimates resulting from individual data, for a given value of N .

In many studies, data collection is limited by the cost of testing rather than the lack of specimens. In this case, a relevant comparison is across a fixed number of tests (i.e., J). In making these comparisons, one will note that all parameter estimates based on pooled data, under HP, are more efficient, when compared to estimates that are obtained from individual-level data. Thus, if a study is limited by the number of tests that can be performed, the process of analyzing biomarker assessments made on pools, rather than individuals, will result in more precise parameter estimation.

Figures 1 and 2 summarize the performance of the proposed covariate-adjusted methods of evaluating biomarkers under HP. In particular, Fig. 1 presents the average estimated length of 95% pointwise confidence intervals for $t_0(X, \theta)$, $YI(X, \theta)$, and $AUC(X, \theta)$ as well as 95% confidence bands for $AUC(X, \theta)$, for all considered values of N . For a fixed value of N , there are no appreciable differences among the average confidence interval (band) lengths across c , which suggests that the process of pooling has little impact on evaluating a biomarker's discriminatory ability. In contrast, for a fixed J , it is apparent that inference based on pooled assessments is more precise, that is, for a fixed value of J the average length of the intervals for $c = 2$ and 4 are roughly 20–30% and 30–50% shorter, respectively, than for $c = 1$.

Figure 2 presents the empirical coverage probabilities for the confidence intervals depicted in Fig. 1. In practically all the considered settings, the confidence intervals for $t_0(X, \theta)$ and $AUC(X, \theta)$ based on both pooled and individual data attained their nominal level, with the few exceptions being explained by the relatively small sample size being analyzed; for example, if $N = 40$ and $c = 4$, then there are only $J = 10$ tests. This phenomenon is more pronounced in the empirical coverage probabilities associated with the confidence intervals for $YI(X, \theta)$, but through other studies we have found that for larger values of N (e.g., $N = 300$), these confidence intervals attain their nominal level. The empirical coverage probabilities for the confidence bands for $AUC(X, \theta)$ were all at their nominal level across all considered configurations, that is, the empirical coverage probabilities when $c = (1, 2, 4)$ were (0.955, 0.951, 0.947) when $N = 40$; (0.952, 0.952, 0.947) when $N = 80$; and (0.951, 0.950, 0.954) when $N = 160$. In summary, these results reinforce the assertion that the process of analyzing pooled, rather than individual, biomarker assessments results in more precise estimation and inference, at the same cost of testing.

The additional simulation studies, which are provided in Supporting Information Appendix C, were designed to assess the performance of the proposed methodology under other scenarios that would likely be encountered in practice. In particular, Supporting Information Simulation Studies 1–3 investigate the effect of violations of the assumption of normality, having a different number of cases and controls, and violations of the assumption that the pool response represents the arithmetic average of the individual biomarker levels, respectively. Briefly, from Supporting Information Simulation Study 1, we have found that the proposed methodology was relatively robust to the violations of the normality assumption for $c \in \{1, 2, 4\}$. Further, the results based on the pooled data were as robust as the results based on the individual-level data. The results from Supporting Information Simulation Study 2 reinforce the main findings discussed above, that is, we found that the proposed methodology can be used for accurately and efficiently assessing the covariate-

adjusted diagnostic efficacy of a biomarker through analyzing measurements taken on pools, even when the number of cases and controls differ. Finally, Supporting Information Simulation Study 3 provides evidence that the proposed methodology can still provide reliable estimation and inference when the arithmetic average assumption is in fact invalid. That is, under both RP and HP, we have found that both minor and extreme violations of this assumption do not appear to bias the regression parameter estimates, while a slight bias, for extreme violations, does appear in the variance estimates.

4 Data application

Recent evidence suggests that inflammation, and more specifically markers of inflammation, has a deleterious effect on many adverse conditions including cancer (Coussens and Werb, 2002), stroke (Chamorro, 2004), and heart disease (Epstein and Ross, 1999; Koenig et al., 1999). IL-6 is a proinflammatory cytokine that has been related to a host of biological functions (Hirano et al., 1986) including coronary heart disease (Yudkin et al., 2000). It is also well known that high levels of cholesterol are associated with coronary heart disease. Our motivating data consist of 40 cases that had recently experienced an MI, and 40 controls. The controls had no previous MIs or other cardiovascular events and were found to be free of coronary heart disease via an electrocardiogram (ECG). The cases consisted of individuals who had an ECG-confirmed MI within the past two weeks and had survived. Cholesterol and IL-6 were measured for all 80 subjects individually; IL-6 was also assessed on pools of size 2 and 4. All 80 individuals contributed to the individual and pooled data. Thus, the individual-level data consist of 80 observations; for pools of size 2, the data consist of 40 observations ($\approx 1/2$ the testing cost); for pools of size 4, the data consist of 20 observations ($\approx 1/4$ the testing cost). Further, since IL-6 measurements were taken on pools as well as on the individual level, this affords us the ability to assess the arithmetic average assumption for this application. In Supporting Information Appendix D, we provide a figure that presents plots of the observed IL-6 measurements taken on pools versus the average of the individual IL-6 measurements for the specimens of which the pool consists, for $c = 2$ and 4. In this figure, it appears that this assumption is reasonable for this application, that is, if this assumption were perfectly met, the plotted points would fall on the identity line, which is provided within each plot.

The preliminary analysis of these data consisted of fitting a first-order linear regression model to the case and control data separately, using cholesterol as the only predictor variable. Estimates of the regression coefficients and error variances were obtained using the techniques outlined in Section 2.1, for both pooled and individual-level data. Basic diagnostic techniques (i.e., residual plots and qq-plots, which are provided in Supporting Information Appendix D), supported the assumptions of normality and constant error variance under all the considered settings. The regression parameter estimates and their associated standard errors for both case and control data, when $c = 1, 2$, and 4, are provided in Table 3. From these results, one notes that the estimates resulting from the individual-level data and pooled data (i.e., when $c = 2$) are largely in agreement, with some discrepancies when $c = 4$. It is important to note that the use of HP was not utilized in the collection of these data. To examine how our methodology would have performed in this setting under HP, we artificially constructed and analyzed HP data using the individual-level

biomarker measurements available in the data set (for further details, see Supporting Information Appendix D). Table 3 also provides estimates of the regression coefficients and error variances under HP for $c \in \{2, 4\}$.

The parameter estimates reported in Table 3 were used for estimating covariate-dependent values of AUC, Youden's index, and the optimal cutpoint. In Fig. 3, we present our estimates of the cholesterol adjusted $t_0(X, \theta)$ (optimal threshold), $YI(X, \theta)$ (Youden index), and $AUC(X, \theta)$ (conditional AUC), for the individual-level data as well as for the data collected on pools of size 2 and 4. The corresponding figure under HP is provided in Supporting Information Appendix D. The figures corresponding to $t_0(X, \theta)$ and $YI(X, \theta)$ also present 95% pointwise confidence intervals, while the figures pertaining to $AUC(X, \theta)$ contain 95% pointwise confidence intervals as well as 95% confidence bands. The point estimates of the covariate-dependent Youden's index and AUC were consistent for the individual, pools of size 2, and pools of size 4 data. The YI and AUC figures showed an increasing trend between the diagnostic accuracy of IL-6 and cholesterol level; that is, for higher cholesterol levels, we estimate that IL-6 is a more accurate predictor of MI status. In fact, for higher levels of cholesterol, the estimated AUC was greater than 0.8 indicating that IL-6 provides good discrimination between MI statuses. Further, for both pooled data sets, we found that $YI(X, \theta) > 0.5$ when the cholesterol level exceeded roughly 250.

For the individual-level data and data collected using pools of size 2, we found that IL-6 is a significant predictor of MI status (i.e., the lower 95% confidence band is greater than 0.5) for subjects with a cholesterol level above approximately 175. In contrast, for the data collected using pools of size 4, IL-6 was found to be a significant predictor of MI status for a subject with a cholesterol level between 200 and 275. As one should expect, the uncertainty in the estimates tends to increase with the pool size. The discrepancies observed in the estimates based on the pooled data when $c = 4$ is likely attributable to the relatively small sample size that was available for analysis (i.e., $J = 10$). This assertion is based on the findings presented in Section 3 corresponding to the use of RP when $c = 4$ and $J = 10$, which is analogous to this portion of the data application. Under these settings, the estimators of the regression parameters and error variance exhibit a great deal of variability, which could explain the discrepancies observed in Table 3 and Fig. 3 for $c = 4$. Although based on our analysis of the artificially constructed HP data in Supporting Information Appendix D, we believe that these discrepancies could have been attenuated if HP had been used for constructing pools in the original study.

5 Discussion

In this paper, we have developed regression methodology for pooled biomarker measurements. This methodology allows one to determine covariate-dependent optimal cutpoints for diagnosis and also an assessment of the efficacy of biomarker diagnoses that depend on independent variables. For example, covariate-adjusted AUC can be used for identifying subpopulations for which a given biomarker has relatively high diagnostic efficacy. Under the assumption of normality, we have provided confidence interval formulae for the optimal cutpoint, Youden's index, and the AUC. Asymptotic confidence bands can be used for assessing the overall efficacy of a biomarker across the entire covariate domain. To

further disseminate this work, software in the form of R code, together with a sample input data set and complete documentation, is available as Supporting Information on the journal's web page.

The finite sample performance of the proposed methods was investigated through simulation. The main conclusions to be made from these simulations are as follows: (i) there is little to no loss in efficiency/accuracy from pooling when the number of individuals is fixed and (ii) when the number of tests (testing cost) is fixed, we observe marked improvement from pooling individuals, especially if pools are created based on covariate information (HP). The asymptotic pointwise confidence intervals for Youden's index tend to be conservative, but 95% confidence intervals and bands for AUC are extremely accurate even when the data arise from testing 10 pools of size 4. Pooling has almost no effect on confidence interval width for a fixed number of individuals. However, when the number of tests is fixed, pooling can significantly reduce confidence interval (band) lengths, that is, through pooling we get much more information for the same testing cost.

In practice, when multiple predictor variables are available, homogeneous pools can be constructed in a similar fashion as was described in Sections 3 and 4. In particular, HP for one predictor variable was accomplished by first ordering the individuals based on the predictor variable and pools were then formed sequentially based on this ordering. Proceeding in this fashion ensures that pools are constructed of individuals with "similar" covariates. In the presence of multiple predictors, clustering algorithms such as *K*-means (Hartigan and Wong, 1979) can be used for identifying individuals with similar covariates, and pools can subsequently be formed from individuals belonging to a common cluster. This again ensures that pools contain individuals with similar covariates. A variant of this technique was investigated in Mitchell et al. (2014), and was shown to be an effective pooling strategy with respect to estimation efficiency.

The advantages of using independent variables to determine cutpoints and for measuring the performance of a diagnostic test are evident in our analysis of the IL-6 MI data. The plots in Fig. 3 indicate that both optimal cutpoint and diagnostic efficacy of IL-6 in determining MI depend on cholesterol level. In terms of assessing the covariate-adjusted discriminatory ability of IL-6, our data analysis found that inference based on the pooled data, when $c = 2$, rendered the same findings as the analysis based on the individual-level data, and at half the cost of data collection. Ultimately, the performance of the proposed methodology is dependent on the pool size used. Consequently, a topic of future research could be centered around identifying the optimal pool size with respect to minimizing the asymptotic variance of the covariate-adjusted measures of diagnostic accuracy when J is held constant. This could be accomplished by using ideas similar to the compound D-optimality criteria, which accommodates different levels of interest in different linear combinations of the unknown parameters.

The IL-6 data considered in Section 4 are unique in the fact that these consist of both pooled and individual-level measurements. This feature allowed the comparison of the estimates based on the pooled data to those that were obtained through analyzing the individual-level information. Additionally, this feature allowed us to assess the arithmetic average

assumption. Our simulation study in Supporting Information indicates that violations of this assumption do not appear to impact the regression parameter estimates, but extreme violations can adversely affect variance estimates. Regretfully, there does not exist formal or informal techniques that can be used for assessing the validity of this assumption, when one has access to pooled data only. Consequently, the development of such techniques could be a beneficial topic for future research, especially since this assumption is common among the associated literature (e.g., see Faraggi et al., 2003; Liu and Schisterman, 2003; Mumford et al., 2006; Bondell et al., 2007; Vexler et al., 2008; Malinovsky et al., 2012).

In our data analysis, the assumption of normality is empirically supported. In other applications, the normality assumption could be suspected. Although we have shown favorable robustness properties (see Supporting Information Appendix C), there is a need to modify the methods in this paper to allow for alternate distributions, for example, distributions that accommodate positive and right-skewed biomarkers. The class of generalized linear models may be helpful in developing methodology under other parametric modeling assumptions, along with estimation and confidence set formulae akin to the formulae provided in this paper. In particular, the primary hurdle with this extension is the development of the regression methodology that can be used for analyzing pooled biomarker assessments under alternate distributional assumptions. The development of such methodology is challenging due to the fact that the distribution of the observed pooled measurement cannot be expressed in closed form under many common models used for analyzing positive right-skewed biomarker levels, for example, the gamma and log-normal distributions. This challenge could be overcome using techniques akin to the techniques presented in Mitchell et al. (2014). Further, the development of these new modeling techniques should account for the effect of limits of detection, which is a feature commonly encountered in biomarker evaluation studies. If this methodology were developed, then the proposed approach could be applied to obtain estimates of the covariate-adjusted optimal cutpoint, Youden's index, and the AUC, by replacing F_+ and F_- by their respective estimates in Section 2.2. Note that it is not expected that analytical expressions similar to the expressions presented in Section 2.3 will exist under other distributional assumptions, thus these estimators would likely have to be obtained via numerical techniques.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Enrique F. Schisterman was supported in part by the Intramural Research Program of the National Institutes of Health *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

References

- Bilder C and Tebbs J (2009). Bias, efficiency, and agreement for group-testing regression models. *Journal of Statistical Computation and Simulation* 79, 67–80. [PubMed: 20046956]
- Bondell D, Liu A, and Schisterman E (2007). Statistical inference based on pooled data: a moment-based estimating equation approach. *Journal of Applied Statistics* 34, 129–140.

- Borges A, Silverberg M, Wentworth D, Grulich A, Fatkenheuer G, Mitsuyasu R, Tambussi G, Sabin C, Neaton J, Lundgren J, and the INSIGHT SMART, ESPRIT, SILCAAT Study Groups. (2013). Predicting risk of cancer during HIV infection: the role of inflammatory and coagulation biomarkers. *AIDS* 27, 1433–1441. [PubMed: 23945504]
- Boyd J, Rifai N, and Annesley T (2012). Statistical methods for test and biomarker evaluation studies: a clinical chemistry series. *Clinical Chemistry* 58, 1273–1274. [PubMed: 22777719]
- Caudill S (2012). Use of pooled samples from the national health and nutrition examination survey. *Statistics in Medicine* 31, 3269–3277. [PubMed: 22492247]
- Chamorro A (2004). Role of inflammation in stroke and atherothrombosis. *Cerebrovascular Diseases* 17 (Suppl. 3), 1–5.
- Coussens L and Werb Z (2002). Inflammation and cancer. *Nature* 420, 860–867. [PubMed: 12490959]
- Dhanasekaran S, Barrette T, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta K, Rubin M, and Chinnaiyan A (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* 412, 822–826. [PubMed: 11518967]
- Dorfman R (1943). The detection of defective members of large populations. *The Annals of Mathematical Statistics* 14, 436–440.
- Epstein F and Ross R (1999). Atherosclerosis an inflammatory disease. *New England Journal of Medicine* 340, 115–126. [PubMed: 9887164]
- Faraggi D, Reiser B, and Schisterman E (2003). ROC curve analysis for biomarkers based on pooled assessments. *Statistics in Medicine* 22, 2515–2527. [PubMed: 12872306]
- Hartigan J and Wong M (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C* 28, 100–108.
- Hirano T, Yasukawa K, Harada H, Taga T, Watanabe Y, Matsuda T, Kashiwamura S, Nakajima K, Koyama K, Iwamatsu A, et al. (1986). Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin. *Nature* 324, 73–76. [PubMed: 3491322]
- Hochberg Y and Tamhane A (1987). *Multiple Comparison Procedures*. John Wiley & Sons, Inc., New York, NY.
- Huang Y, Fong Y, Wei J, and Feng Z (2011). Borrowing information across populations in estimating positive and negative predictive values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60, 633–653.
- Janes H and Pepe M (2008). Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *American Journal of Epidemiology* 168, 89–97. [PubMed: 18477651]
- Janes H and Pepe M (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* 96, 371–382. [PubMed: 22822245]
- Kanekar A (2010). Biomarkers predicting progression of human immunodeficiency virus-related disease. *Journal of Clinical Medicine Research* 2, 55–61. [PubMed: 21811520]
- Koenig W, Sund M, Frohlich M, Fischer H, Lowel H, Doring A, Hutchinson W, and Pepys M (1999). C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men results from the MONICA (monitoring trends and determinants in cardiovascular disease) Augsburg cohort study, 1984 to 1992. *Circulation* 99, 237–242. [PubMed: 9892589]
- Lau B, Gange S, Kirk G, and Moore R (2009). Evaluation of human immunodeficiency virus biomarkers: inferences from interval and clinical cohort studies. *Epidemiology* 20, 664–672. [PubMed: 19478669]
- Liu A and Schisterman E (2003). Comparison of diagnostic accuracy of biomarkers with pooled assessments. *Biometrical Journal* 45, 631–644.
- Liu A, Schisterman E, and Teoh E (2004). Sample size and power calculation in comparing diagnostic accuracy of biomarkers with pooled assessments. *Journal of Applied Statistics* 31, 49–59.
- Malinovsky Y, Albert P, and Schisterman E (2012). Pooling designs for outcomes under a Gaussian random effects model. *Biometrics* 68, 45–52. [PubMed: 21981372]

- Mitchell E, Lyles R, Manatunga A, Danaher M, Perkins N, and Schisterman E (2014). Regression for skewed biomarker outcomes subject to pooling. *Biometrics* 70, 202–211. [PubMed: 24521420]
- Mumford S, Schisterman E, Vexler A, and Liu A (2006). Pooling biospecimens and limits of detection: effects on ROC curve analysis. *Biostatistics* 7, 585–598. [PubMed: 16531470]
- Rasmussen T, Tolstrup M, Melchjorsen J, Frederiksen C, Nielsen U, Langdahl B, Østergaard L, and Laursen A (2011). Evaluation of cardiovascular biomarkers in HIV-infected patients switching to abacavir or tenofovir based therapy. *BMC Infectious Diseases* 11, 267. [PubMed: 21970555]
- Sabatine M, Morrow D, de Lemos J, Omland T, Sloan S, Jarolim P, Solomon S, Pfeffer M, and Braunwald E (2012). Evaluation of multiple biomarkers of cardiovascular stress for risk prediction and guiding medical therapy in patients with stable coronary disease. *Circulation* 125, 233–240. [PubMed: 22179538]
- Schisterman E, Faraggi D, and Reiser B (2004). Adjusting the generalized roc curve for covariates. *Statistics in Medicine* 23, 3319–3331. [PubMed: 15490426]
- Schisterman E, Perkins N, Liu A, and Bondell H (2005). Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology* 16, 73–81. [PubMed: 15613948]
- Vansteelandt S, Goetghebeur E, and Verstraeten T (2000). Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics* 56, 1126–1133. [PubMed: 11129470]
- Vexler A, Schisterman E, and Liu A (2008). Estimation of ROC curves based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Statistics in Medicine* 27, 280–296. [PubMed: 17721905]
- Wang D, McMahan C, Gallagher C, and Kulasekera K (2014) Semiparametric group testing regression models. *Biometrika* 101, 587–598.
- Youden W (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35. [PubMed: 15405679]
- Yudkin J, Kumari M, Humphries S, and Mohamed-Ali V (2000). Inflammation, obesity, stress and coronary heart disease: is interleukin-6 the link? *Atherosclerosis* 148, 209–214. [PubMed: 10657556]
- Zhang B, Bilder C, and Tebbs J (2013). Group testing regression model estimation when case identification is a goal. *Biometrical Journal* 55, 173–189. [PubMed: 23401252]

Figure 1.

Simulation results under HP. The top, middle, and bottom rows of the panels present the average length of 95% pointwise confidence intervals for $t_0(X, \theta)$, $YI(X, \theta)$, and $AUC(X, \theta)$, respectively, across the predictor space and for all considered values of N . The panels on the left, center, and right correspond to $N=40$, 80 , and 160 , respectively. Also provided are the average lengths of the confidence bands for $AUC(X, \theta)$ (bottom row, gray).

Figure 2.

Simulation results under HP: The top, middle, and bottom rows of the panels present the empirical coverage probabilities associated with 95% pointwise confidence intervals for $t_0(X, \theta)$, $YI(X, \theta)$, and $AUC(X, \theta)$, respectively, across the predictor space and for all considered values of N . In particular the panels on the left, center, and right correspond to $N = 40, 80$, and 160 , respectively.

Figure 3.

Interleukin-6 data analysis: Presented results include the covariate-dependent estimates of $t_0(\mathbf{X}, \theta)$ (left panels), $YI(\mathbf{X}, \theta)$ (center panels), and $AUC(\mathbf{X}, \theta)$ (right panels). Also included are 95% pointwise confidence intervals (black) for $t_0(\mathbf{X}, \theta)$, $YI(\mathbf{X}, \theta)$, and $AUC(\mathbf{X}, \theta)$, as well as 95% confidence bands (gray) for $AUC(\mathbf{X}, \theta)$. The dot-dash line provided in the plots pertaining to $AUC(\mathbf{X}, \theta)$, can be used for identifying the range of cholesterol levels for which IL-6 is an accurate predictor of MI status.

Table 1

Simulation results under HP: Presented results include the mean (Mean) and standard deviation (SD) of the parameter estimates, as well as average estimated standard errors (SE) and estimated coverage probabilities (Cov) for 95% confidence intervals.

$N = 40$		$c = 1 \ (J = 40)$	$c = 2 \ (J = 20)$	$c = 4 \ (J = 10)$
$\beta_{1+} = 1.750$	Mean (SD) SE (Cov)	1.756(1.994) 2.019 (0.950)	1.760 (2.005) 2.029 (0.950)	1.760 (2.038) 2.054 (0.950)
$\beta_{2+} = 0.015$	Mean (SD) SE (Cov)	0.015 (0.009) 0.009 (0.953)	0.015 (0.009) 0.009 (0.950)	0.015 (0.009) 0.009 (0.950)
$\beta_{1-} = 3.000$	Mean (SD) SE (Cov)	2.998 (1.165) 1.158 (0.958)	2.998 (1.173) 1.164(0.959)	3.002(1.187) 1.180 (0.958)
$\beta_{2-} = -0.005$	Mean (SD) SE (Cov)	-0.005 (0.006) 0.006 (0.943)	-0.005 (0.006) 0.006 (0.940)	-0.005 (0.006) 0.006 (0.943)
$\sigma_+^2 = 4.623$	Mean (SD) SE (Cov)	4.630 (1.058) 1.062(0.953)	4.623 (1.532) 1.540(0.952)	4.617(2.313) 2.309 (0.949)
$\sigma_-^2 = 1.823$	Mean (SD) SE (Cov)	1.829 (0.419) 0.419 (0.955)	1.824(0.610) 0.608 (0.947)	1.829 (0.936) 0.919 (0.946)
$N = 80$		$c = 1 \ (J = 80)$	$c = 2 \ (J = 40)$	$c = 4 \ (J = 20)$
$\beta_{1+} = 1.750$	Mean (SD) SE (Cov)	1.760(1.409) 1.398 (0.948)	1.761 (1.414) 1.403 (0.948)	1.760(1.423) 1.412(0.950)
$\beta_{2+} = 0.015$	Mean (SD) SE (Cov)	0.015 (0.006) 0.006 (0.947)	0.015 (0.006) 0.006 (0.948)	0.015 (0.006) 0.006 (0.949)
$\beta_{1-} = 3.000$	Mean (SD) SE (Cov)	3.016(0.791) 0.805 (0.965)	3.016(0.793) 0.806 (0.965)	3.015 (0.798) 0.810 (0.965)
$\beta_{2-} = -0.005$	Mean (SD) SE (Cov)	-0.005 (0.004) 0.004 (0.947)	-0.005 (0.004) 0.004 (0.947)	-0.005 (0.004) 0.004(0.951)
$\sigma_+^2 = 4.623$	Mean (SD) SE (Cov)	4.620 (0.739) 0.740 (0.952)	4.632(1.064) 1.063 (0.949)	4.645 (1.539) 1.547 (0.952)
$\sigma_-^2 = 1.823$	Mean (SD) SE (Cov)	1.828 (0.295) 0.293 (0.947)	1.825 (0.420) 0.419 (0.948)	1.826(0.608) 0.609 (0.949)
$N = 160$		$c = 1 \ (J = 160)$	$c = 2 \ (J = 80)$	$c = 4 \ (J = 40)$
$\beta_{1+} = 1.750$	Mean (SD) SE (Cov)	1.751 (0.986) 0.981 (0.948)	1.752(0.987) 0.981 (0.949)	1.752(0.989) 0.984 (0.949)
$\beta_{2+} = 0.015$	Mean (SD) SE (Cov)	0.015 (0.004) 0.004 (0.947)	0.015 (0.004) 0.004 (0.946)	0.015 (0.004) 0.004 (0.950)
$\beta_{1-} = 3.000$	Mean (SD) SE (Cov)	3.003 (0.560) 0.563 (0.966)	3.003 (0.560) 0.563 (0.967)	3.003 (0.562) 0.565 (0.965)
$\beta_{2-} = -0.005$	Mean (SD) SE (Cov)	-0.005 (0.003) 0.003 (0.950)	-0.005 (0.003) 0.003 (0.949)	-0.005 (0.003) 0.003 (0.948)
$\sigma_+^2 = 4.623$	Mean (SD) SE (Cov)	4.630 (0.521) 0.521 (0.950)	4.625 (0.744) 0.741 (0.953)	4.632(1.068) 1.063 (0.950)
$\sigma_-^2 = 1.823$	Mean (SD) SE (Cov)	1.825 (0.205) 0.205 (0.949)	1.824(0.289) 0.292 (0.948)	1.826(0.417) 0.419 (0.952)

Table 2

Simulation results under RP: Presented results include the mean (Mean) and standard deviation (SD) of the parameter estimates, as well as average estimated standard errors (SE) and estimated coverage probabilities (Cov) for 95% confidence intervals.

$N = 40$		$c = 1$ ($J = 40$)	$c = 2$ ($J = 20$)	$c = 4$ ($J = 10$)
$\beta_{I+} = 1.750$	Mean (SD)	1.771 (2.018)	1.797 (2.966)	1.788 (4.564)
	SE (Cov)	2.019 (0.950)	2.954 (0.949)	4.566 (0.950)
$\beta_{2+} = 0.015$	Mean (SD)	0.015 (0.009)	0.015 (0.013)	0.015 (0.020)
	SE (Cov)	0.009 (0.953)	0.013 (0.950)	0.020 (0.950)
$\beta_{I-} = 3.000$	Mean (SD)	2.975 (1.160)	2.973 (1.691)	2.959 (2.623)
	SE (Cov)	1.158 (0.958)	1.688 (0.954)	2.624 (0.943)
$\beta_{2-} = -0.005$	Mean (SD)	-0.005 (0.006)	-0.005 (0.008)	-0.005 (0.013)
	SE (Cov)	0.006 (0.943)	0.008 (0.936)	0.013 (0.925)
$\sigma_+^2 = 4.623$	Mean (SD)	4.628 (1.051)	4.617 (1.520)	4.615 (2.310)
	SE (Cov)	1.061 (0.953)	1.537 (0.952)	2.308 (0.950)
$\sigma_-^2 = 1.823$	Mean (SD)	1.829 (0.418)	1.824 (0.608)	1.829 (0.930)
	SE (Cov)	0.419 (0.952)	0.608 (0.948)	0.918 (0.947)
$N = 80$		$c = 1$ ($J = 80$)	$c = 2$ ($J = 40$)	$c = 4$ ($J = 20$)
$\beta_{I+} = 1.750$	Mean (SD)	1.763 (1.376)	1.757 (1.995)	1.731 (2.900)
	SE (Cov)	1.398 (0.953)	2.006 (0.952)	2.958 (0.954)
$\beta_{2+} = 0.015$	Mean (SD)	0.015 (0.006)	0.015 (0.009)	0.015 (0.013)
	SE (Cov)	0.006 (0.952)	0.009 (0.952)	0.013 (0.953)
$\beta_{I-} = 3.000$	Mean (SD)	2.996 (0.806)	3.000 (1.140)	2.987 (1.668)
	SE (Cov)	0.804 (0.962)	1.149 (0.960)	1.682 (0.953)
$\beta_{2-} = -0.005$	Mean (SD)	-0.005 (0.004)	-0.005 (0.006)	-0.005 (0.008)
	SE (Cov)	0.004 (0.947)	0.006 (0.942)	0.008 (0.936)
$\sigma_+^2 = 4.623$	Mean (SD)	4.623 (0.740)	4.636 (1.062)	4.658 (1.542)
	SE (Cov)	0.740 (0.952)	1.063 (0.950)	1.552 (0.951)
$\sigma_-^2 = 1.823$	Mean (SD)	1.827 (0.294)	1.824 (0.419)	1.825 (0.608)
	SE (Cov)	0.293 (0.948)	0.419 (0.947)	0.608 (0.948)
$N = 160$		$c = 1$ ($J = 160$)	$c = 2$ ($J = 80$)	$c = 4$ ($J = 40$)
$\beta_{I+} = 1.750$	Mean (SD)	1.733 (0.974)	1.727 (1.379)	1.709 (1.997)
	SE (Cov)	0.981 (0.952)	1.389 (0.951)	1.999 (0.950)
$\beta_{2+} = 0.015$	Mean (SD)	0.015 (0.004)	0.015 (0.006)	0.015 (0.009)
	SE (Cov)	0.004 (0.955)	0.006 (0.951)	0.009 (0.950)
$\beta_{I-} = 3.000$	Mean (SD)	3.001 (0.561)	2.992 (0.784)	3.001 (1.143)
	SE (Cov)	0.563 (0.966)	0.796 (0.966)	1.142 (0.962)
$\beta_{2-} = -0.005$	Mean (SD)	-0.005 (0.003)	-0.005 (0.004)	-0.005 (0.006)
	SE (Cov)	0.003 (0.949)	0.004 (0.947)	0.006 (0.942)
$\sigma_+^2 = 4.623$	Mean (SD)	4.631 (0.521)	4.627 (0.744)	4.635 (1.069)
	SE (Cov)	0.521 (0.950)	0.741 (0.954)	1.064 (0.950)
$\sigma_-^2 = 1.823$	Mean (SD)	1.825 (0.205)	1.825 (0.289)	1.826 (0.416)
	SE (Cov)	0.205 (0.951)	0.292 (0.948)	0.419 (0.953)

Table 3

Interleukin data analysis: Presented results include the regression parameter estimates (estimated standard errors) for both case and control models, when $c = 1, 2$, and 4 , obtained through analyzing the observed pooled IL-6 measurements available in the data and the artificially constructed pooled IL-6 measurements that resulted from HP.

			$c = 1 (J = 40)$	$c = 2 (J = 20)$	$c = 4 (J = 10)$
Original data	Case	Intercept	1.717 (1.904)	1.974 (2.478)	0.570 (5.442)
		Cholesterol	0.011 (0.008)	0.010 (0.010)	0.017 (0.024)
	Control	Intercept	3.029 (1.081)	2.399 (1.397)	4.156(2.185)
		Cholesterol	-0.006 (0.005)	-0.003 (0.008)	-0.012(0.011)
HP data	Case	Intercept	1.717(1.904)	1.566 (2.141)	1.756(2.620)
		Cholesterol	0.011 (0.008)	0.012(0.009)	0.011 (0.011)
	Control	Intercept	3.029 (1.081)	2.809(1.148)	2.732(1.329)
		Cholesterol	-0.006 (0.005)	-0.004 (0.005)	-0.004 (0.006)