Title

A mixed effects model for analyzing area under the curve of longitudinally measured biomarkers with missing data

Authors

Shi, Luoxi
Hatsukami, Dorothy K
Koopmeiners, Joseph S
et al.

Peer reviewed

A mixed effects model for analyzing area under the curve of longitudinally measured biomarkers with missing data

Luoxi Shi[1], **Dorothy K. Hatsukami**[2,3], **Joseph S. Koopmeiners**[1], **Chap T. Le**[1], **Neal L. Benowitz**[4], **Eric C. Donny**[5], **Xianghua Luo**[1,2]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA

[2]Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, USA

[3]Department of Psychiatry, Medical School, University of Minnesota, Minneapolis, Minnesota, USA

[4]Department of Medicine, University of California San Francisco, San Francisco, California, USA

[5]Baptist Comprehensive Cancer Center and Department of Physiology and Pharmacology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

Abstract

A simple approach for analyzing longitudinally measured biomarkers is to calculate summary measures such as the area under the curve (AUC) for each individual and then compare the mean AUC between treatment groups using methods such as *t* test. This two-step approach is difficult to implement when there are missing data since the AUC cannot be directly calculated for individuals with missing measurements. Simple methods for dealing with missing data include the complete case analysis and imputation. A recent study showed that the estimated mean AUC difference between treatment groups based on the linear mixed model (LMM), rather than on individually calculated AUCs by simple imputation, has negligible bias under random missing assumptions and only small bias when missing is not at random. However, this model assumes the outcome to be normally distributed, which is often violated in biomarker data. In this paper, we propose to use a LMM on log-transformed biomarkers, based on which statistical inference for the ratio, rather than difference, of AUC between treatment groups is provided. The proposed method can not only handle the potential baseline imbalance in a randomized trail but also circumvent the estimation of the nuisance variance parameters in the log-normal model. The proposed model is applied to a

recently completed large randomized trial studying the effect of nicotine reduction on biomarker exposure of smokers.

Keywords

area under the curve; biomarker; longitudinal; missing data; mixed effects model

1 | INTRODUCTION

Cigarette smoking is responsible for most of tobacco-related mortality and morbidity.[1] Reducing the nicotine content of cigarettes is a strategy to reduce the addictiveness of cigarettes,[2,3] which can also reduce the prevalence of smoking and result in the reduction of tobacco toxicant exposures.[4] A large-scale randomized trial studying the effects of immediate vs gradual nicotine reduction on tobacco toxicant exposures was recently conducted by the Center for the Evaluation of Nicotine in Cigarettes (CENIC).[5] The primary outcome variables of this study were a panel of biomarkers repeatedly measured over a 20-week intervention period.

In longitudinal studies, a simple approach to analyze repeatedly measured data is to calculate a summary measure for each individual.[6] The area under the measurement-time curve (AUC) is one commonly used summary measure. Besides the statistical simplicity, the AUC also has scientific or clinical significance that makes it useful for various types of research.[7-9] In the CENIC study, the intent was to determine the cumulative exposure to cigarette smoke toxicants over time across different approaches to reducing nicotine in cigarettes, referred in this article as treatment groups, and hence the AUC of the concentration of toxicant biomarkers was chosen as the primary endpoint.[5]

In the absence of missing data, a common approach to estimate AUC is using the linear trapezoidal method.[10] After obtaining the AUC for each individual, the average AUC between treatment groups can be analyzed using simple statistical methods such as the *t* test or the nonparametric Wilcoxon rank-sum test. However, missing data are common in longitudinal studies.[11,12] Some frequently encountered missing data patterns in longitudinal studies include dropout and intermittent missing. Missingness can be completely random (called missing completely at random or MCAR), dependent on the observed data (called missing at random or MAR), or dependent on both the observed and unobserved data (called missing not at random or MNAR). A comprehensive review of missing data methods can be found in Little and Rubin.[13]

In the presence of missing data, a simple two-step approach can be used to compare AUCs between different groups: first calculating each individual's AUC using methods such as complete case, simple imputation, and multiple imputation (MI), then comparing the mean AUC between different groups as adopted in the planned primary analysis of the CENIC data which motivated our study.[5] It has been shown that tests based on estimated AUC using a simple imputation may suffer some poor properties including inflated type I error and poor power, even when the data are MCAR.[14] Alternatively, one can fit a linear mixed model[12] or a general linear model[7] on the repeatedly measured biomarkers and construct

the difference in AUC between groups using the estimated regression parameters. The linear mixed model has been shown to perform better than using individually calculated AUC with imputed data.[12] Although these model-based methods provide a way to analyze AUCs with missing data in repeatedly measured biomarkers, there are some issues that remain to be solved. First, in the linear mixed model, the normal distribution is assumed for the error term and random effects, which implies that the linear outcome variable also follows a normal distribution. However, many biomarkers are found right-skewed,[15,16] including those studied in tobacco research.[17] And in such circumstances, the log-normal distribution has been found more appropriate than the normal distribution for biomarkers.[18] To normalize such variables, a log-transformation on the response variable before the analysis is needed and the relative effect rather than absolute effect of predictors can be estimated from such models.[19] Second, the existing mixed model for analyzing the AUC[12] does not consider the possible imbalance in the biomarker between treatment groups at baseline, while by chance unforeseen imbalances in baseline measurements may be observed in a randomized trial.[20] Under this circumstance, one may adjust for baseline measurements using an analysis of covariance (ANCOVA) approach as suggested by a guidance document for analyzing randomized trials by the Food and Drug Administration.[21] This method has been adopted in the published CENIC study[5] and other studies.[22]

In this paper, we propose a mixed effects model for log-normally distributed biomarker outcomes, based on which, statistical inference for the comparison of AUC between treatment groups can be provided. In addition, we propose to use the baseline-scaled AUC when calculating the AUC ratio between two groups, to adjust for the potential imbalance in baseline biomarker levels. We show that by using the baseline-scaled AUC we can not only handle the potential baseline imbalance in a randomized trial but also circumvent the estimation of nuisance variance parameters in the log-normal model. Monte-Carlo simulations were conducted to compare the performance of the proposed mixed effects model to the two-step approach with two frequently used imputation approaches in tobacco research: the MI approach and the last observation carried forward (LOCF) simple imputation, while we acknowledge that the LOCF is not recommended as the primary approach in applications due to its bias, unless its assumption can be justified scientifically.[23,24] The proposed mixed effects model was then applied to the CENIC data to analyze the effects of different approaches to nicotine reduction.

2 | MOTIVATING DATA EXAMPLE

The CENIC study was a randomized, multicenter, parallel and double-blinded study, including 1250 subjects who were randomized to one of the three groups: (a) immediate reduction to very low nicotine content; (b) gradual reductions of nicotine content in cigarettes; (c) conventional levels of nicotine in cigarettes (control). The primary outcome variables included: expired carbon monoxide (CO), urine 3-hydroxypropylmercapturic (3-HPMA), a metabolite of the acrolein, a suspected cardiopulmonary toxin, and urinary phenanthrene tetraol (PheT), which is a biomarker of exposure to polycyclic aromatic hydrocarbons and metabolic activation of this class of carcinogens. In this paper, we focused on the biomarkers 3-HPMA and PheT, which were shown to be right-skewed. These biomarkers were measured every 4 weeks, including the baseline visit, over a 20-week

intervention period. The primary outcome of interest in this study was the area under the biomarker concentration-time curve or AUC, as a measure of the cumulative toxicant exposure of smokers. More details of the design and the population of this trial can be found in the main outcome paper.[5]

Figure 1 shows the trajectories of 3-HPMA and PheT over time for different treatment groups. A *t* test showed the potential imbalance in baseline PheT (on the log-transformed values). The missing data of 3-HPMA and PheT are summarized by visits and by different missing patterns in Table 1. The quantile-quantile (Q-Q) plots of the two biomarkers and their log-transformed values at baseline (Figure A1 in Appendix A) showed that the two biomarkers are skewed and log-transformation is appropriate for the two biomarkers.

3 | METHODS

In this study, we proposed a mixed effects model for log-normally distributed biomarker outcomes. Let $j = 0, 1,\ldots, m$ index visits, and $g = 0, 1,\ldots, k$ index treatment groups with $g = 0$ indicating the control group. We considered a balanced design as the CENIC study, where all subjects in the study had the same scheduled visits at time $t_0, t_1, t_2,\ldots, t_m$, where $t_0 = 0$ indicates baseline, and t_j indicates the time interval between two adjacent visits t_{j-1} and t_j, for $j \geq 1$. Note that in the CENIC study, $t_j = 4$ weeks for all j. The mixed model which allows a flexible treatment effect pattern over time includes the discrete time variable (visit), treatment group, and their interaction, in the following form:

$$\log(Y_{ij}) = \beta_0 + \beta_j^J + \sum_{g=1}^{k} \beta_g^G I(G_i = g) + \sum_{g=1}^{k} \beta_{gj}^{GJ} I(G_i = g) + b_i + e_{ij} \text{ for } j = 1,\ldots, m, \text{ and}$$

$$\log(Y_{i0}) = \beta_0 + \sum_{g=1}^{k} \beta_g^G I(G_i = g) + b_i + e_{i0} \text{ for } j = 0,$$

(1)

where Y_{ij} is the biomarker value of the *i*th participant at the *j*th visit, G_i denotes the treatment group of the participant, $b_i \sim N(0, \sigma_b^2)$ represents between-person effects, and $e_{ij} \sim N(0, \sigma_e^2)$ represents the random error. The regression coefficient β_0 is intercept and the mean log-biomarker of the control group at baseline, while the superscripts J, G, and GJ for β are for distinguishing the effect of time (discrete), group, and the interaction of time and group, and the subscripts j, g, and gj are for different strata of these factors. Specifically, $\beta_j^J, j = 1,\ldots, m$ are the differences of the mean log-biomarker of the control group at visit j compared with the baseline; $\beta_g^G, g = 1,\ldots, k$ are the differences of the mean log-biomarker of group g at baseline compared with the control group's baseline; and $\beta_{gj}^{GJ}, j = 1,\ldots, m, g = 1,\ldots, k$ are the regression coefficients for the interactions between visit and treatment, which are the difference between treatment group g and control in terms of the change in log-biomarker at visit j from baseline. We denote the estimated regression coefficients by $\hat{\beta}$'s and the estimated variance parameters by $\hat{\sigma}^2$'s.

We know that for a log-normally distributed variable Y, with $\log Y \sim N(\mu, \sigma_Y^2)$, the expectation of Y is $E(Y) = \exp(\mu + \sigma_Y^2 / 2)$, rather than the simple transformation of the mean of $\log Y$, $\exp(\mu)$. Thus, the mean biomarker level of group g at t_j can be estimated by:

$$\hat{\mu}_{gj} = \exp\left\{\hat{\beta}_0 + \hat{\beta}_j^J + \hat{\beta}_g^G + \hat{\beta}_{gj}^{GJ} + \left(\hat{\sigma}_b^2 + \hat{\sigma}_e^2\right)/2\right\} \text{for } j = 1, \ldots, m, \text{ and}$$
$$\hat{\mu}_{g0} = \exp\left\{\hat{\beta}_0 + \hat{\beta}_g^G + \left(\hat{\sigma}_b^2 + \hat{\sigma}_e^2\right)/2\right\} \text{for } j = 0. \tag{2}$$

Hence, the AUC for group g based on the estimated mean biomarker level at each visit can be estimated by:

$$\begin{aligned}
\widehat{AUC}_g &= \sum_{j=1}^m \Delta t_j\left(\hat{\mu}_{g,j-1} + \hat{\mu}_{gj}\right)/2 \\
&= \frac{1}{2}\exp\left\{\hat{\beta}_0 + \hat{\beta}_g^G + \left(\hat{\sigma}_b^2 + \hat{\sigma}_e^2\right)/2\right\} \\
&\quad \times \left[t_1\left\{1 + \exp\left(\hat{\beta}_1^J + \hat{\beta}_{g1}^{GJ}\right)\right\} + \sum_{j=2}^m \Delta t_j \right. \\
&\quad \left. \left\{\exp\left(\hat{\beta}_{j-1}^J + \hat{\beta}_{g,j-1}^{GJ}\right) + \exp\left(\hat{\beta}_j^J + \hat{\beta}_{gj}^{GJ}\right)\right\}\right].
\end{aligned} \tag{3}$$

We propose a baseline-scaled AUC to deal with the potential imbalance of baseline biomarkers between groups as follows:

$$A\widetilde{U}C_g = A\widehat{U}C_g / A\widehat{U}C_{g0}, \tag{4}$$

where $A\widehat{U}C_{g0} = t_m\hat{\mu}_{g0} = t_m\exp\left\{\hat{\beta}_0 + \hat{\beta}_g^G + \left(\hat{\sigma}_b^2 + \hat{\sigma}_e^2\right)/2\right\}$ is the estimated mean AUC of group g if the biomarker level were time-invariant over the whole time period, 0 to t_m. The baseline-scaled AUC in (4) can then be expressed as:

$$\begin{aligned}
A\widetilde{U}C_g &= \frac{1}{2t_m}\left[t_1\left\{1 + \exp\left(\hat{\beta}_1^J + \hat{\beta}_{g1}^{GJ}\right)\right\} + \sum_{j=2}^m \Delta t_j \right. \\
&\quad \left. \left\{\exp\left(\hat{\beta}_{j-1}^J + \hat{\beta}_{g,j-1}^{GJ}\right) + \exp\left(\hat{\beta}_j^J + \hat{\beta}_{gj}^{GJ}\right)\right\}\right].
\end{aligned} \tag{5}$$

It is noteworthy to mention that the AUC estimator for group g in Equation (3) and that for the control group both involve the estimators for the two variance parameters σ_b^2 and σ_e^2, which are usually considered as nuisance parameters, while the baseline-scaled AUC estimator in Equation (5) is free of the nuisance parameters. Additionally, the baseline-scaled AUC estimator in Equation (5) is free of the baseline parameters β_0 and β_g^G.

One can then calculate the ratio of baseline-scaled AUCs between different groups g and g' as $A\widetilde{U}C_g / A\widetilde{U}C_{g'}$.

The maximum likelihood estimation (MLE) method can be used for the estimation of the parameters in the mixed model. The variance of the log-transformed ratio of baseline-scaled AUCs can be estimated using the delta method, which can then be used to construct 95% confidence intervals (CIs) for the log-transformed ratio of baseline-scaled AUCs and the ratio of baseline-scaled AUCs after a simple exponentiation transformation.

4 | SIMULATIONS

4.1 | Simulation methods

We conducted a series of Monte-Carlo simulation studies with scenarios mimicking the CENIC data except that only two treatment groups were generated, that is, $g = 0, 1$. Different trajectories of the biomarkers, missing data patterns, missing data mechanisms, and missing rates were considered. For each scenario, a total of 1000 Monte Carlo simulations were performed, with 200 subjects (100 treatment and 100 control) per simulation. Six regular visits, including the baseline visit ($j = 0, \ldots, 5$) were simulated for each subject. For simplicity, we assumed that the mean biomarker level for the control group did not change over time, that is, $\beta_j^J = 0$, and all subjects had the same scheduled, equally spaced visits with $t_j = 1$ for $j \geq 1$.

First, to examine the type I error rate for different methods, we simulated data where there was no difference in the baseline-adjusted AUC between the two groups (ie, no treatment effect) while allowing the baseline biomarker level of the two groups to be unbalanced, by setting $\beta = (\beta_0, \beta_1^G, \beta_{1,1}^{GJ}, \beta_{1,2}^{GJ}, \beta_{1,3}^{GJ}, \beta_{1,4}^{GJ}, \beta_{1,5}^{GJ}) = (2, 1, 0, 0, 0, 0, 0)$, where $\beta_1^G \neq 0$ indicates that the biomarker is unbalanced at baseline between the two groups. Under this scenario, the true value of baseline-scaled AUC ratio = 1.

We then considered two simulation scenarios with nonzero treatment effect but with an unbalanced baseline (by setting $\beta_1^G \neq 0$) (See Figure 2A,B). Specifically, in Panel A, the treatment group had a linear decline trend in the expectation of $\log(Y_{ij})$ by setting $\beta = (2, 1, -0.08, -0.16, -0.24, -0.32, -0.40)$. Under this scenario, the true value of log(baseline-scaled AUC ratio) = −0.19, implying a positive treatment effect if the biomarker is harmful. The scenario in Panel B is similar to that in Panel A, except that the treatment group had a linear spline decline trend in the expectation of $\log(Y_{ij})$ by setting $\beta = (2, 1, -0.10, -0.20, -0.30, -0.30, -0.30)$ (Figure 2B). Under this scenario, the true value of log(baseline-scaled AUC ratio) = −0.20.

The performance of the proposed model when the baseline biomarker is balanced was investigated with additional simulation scenarios by setting $\beta_1^G = 0$ but keeping the other parameters (the other β's and all p_j's) the same as previously assumed. The trajectories of the biomarker for these scenarios are presented in the lower panel of Figure 2C,D.

After simulating the "complete" data following the model described above, we generated various missing data patterns with different missing rates (25% and 50%, referred to as "low" and "high" missing rate, respectively) under different missing data mechanisms. Two missing data patterns were considered: dropout and intermittent missing data, and two

missing data mechanisms, MCAR and MAR were considered. To generate MCAR data, we randomly deleted observations according to the missing pattern based on missing rates. For dropout, none of the baseline observations are missing, and the probability of dropout for the other visit is $p_j = 1 - (1 - \text{missing rate})^{1/5}$, $j = 1,2,\ldots,5$; for intermittent missing data, neither the baseline observations or the last observations are missing, the probability of missing for the other visits is $p_j = 1 - (1 - \text{missing rate})^{1/4}$, $j = 1,2,\ldots,4$. When generating MAR data, the missing probabilities depended on the observed response values, and the probability that Y_{ij} ($j = 1,\ldots,5$ for dropout missing, $j = 1,\ldots,4$ for intermittent missing) was set to missing was a function of $\log(Y_{i,j-1})$ and $\log(\bar{Y}_{j-1})$, where $\bar{Y}_{j-1}$ is the mean of the observed responses up to time t_{j-1}:

$$1 / \left(1 + \exp\left[-\left\{\log\left(Y_{i,\,j-1}\right) / \log\left(\bar{Y}_{j-1}\right)\right\} \log\left\{p_j / \left(1 - p_j\right)\right\}\right]\right).$$

When simulating MAR dropout with low missing data, we set $p_j = 0.018, 0.028, 0.038, 0.048, 0.058$ for $j = 1,\ldots,5$, respectively. When simulating MAR dropout with high missing data, we set $p_j = 0.078, 0.098, 0.118, 0.138, 0.158$ for $j = 1,\ldots,5$, respectively. When simulating MAR intermittent missing data with low missing rate, we set $p_j = 0.0355, 0.0455, 0.0555, 0.0655$ for $j = 1,\ldots,4$, respectively. When simulating MAR intermittent missing with high missing rate, we set $p_j = 0.12, 0.14, 0.16, 0.18$ for $j = 1,\ldots,4$, respectively.

We compared the performance of the proposed mixed effects model with the two-step methods using the LOCF simple imputation and the MI methods. Specifically, for the LOCF method, we carried forward the last available observation for dropout or intermittent missing and carried backward the first available observation for baseline missing. We used the Markov chain Monte Carlo (MCMC) method for MI[11,25] with 20 imputations, where variables were log-transformed before imputation and imputation was performed separately for each treatment group. The AUC for each individual was then calculated based on the imputed data and log-transformed as the outcome variable in a linear regression: $\log\{AUC_i/(t_m Y_{i0})\}$, where the denominator $t_m Y_{i0}$ is the area under the straight line at the baseline biomarker level Y_{i0}, which renders the ratio to be the individual-level analogy to $A\widetilde{U}C$, the baseline-scaled AUC at the population level in the proposed method. We evaluated the performance of the two-step methods and the proposed mixed effects model by comparing their coverage, power, relative bias in estimating the baseline-scaled AUC ratio, and the Monte-Carlo empirical standard deviation (SD) and the mean standard error (SE) of the log-transformed baseline-scaled AUC ratio.

4.2 | Simulation results

When there was no treatment effect, that is, no difference between the treatment and control groups in terms of the baseline-adjusted AUC, the type I error rate of the proposed mixed effects model and the comparison methods were all close to 0.05, regardless of the extent of missing data or missing data mechanism (not shown).

The two simulation scenarios with nonzero treatment effect with unbalanced baseline are presented in Table 2 for the scenario when the treatment group's log-biomarker has a linear decline trend or a linear spline trend. When there were no missing data (shown in the top

panel of Table 2), the estimation of both individually calculated AUC (with no imputation) and the proposed mixed effects model showed negligible bias, as expected. However, in the presence of missing data, the LOCF estimates were biased in all scenarios and the bias increased with the extent of missing data, even when the missing data mechanism was MCAR. The MI with MCMC method and the proposed mixed effects model showed negligible bias under all scenarios, while the proposed mixed effects model had smaller SEs and larger power than the MI method under all simulation scenarios.

The results when the baseline biomarker is balanced for different missing rates and missing patterns are presented in Table 3. Note that when there are no missing data or the missing data mechanism was MCAR, the results of the simulation scenarios with balanced baseline (the top panel of Table 3) or unbalanced baseline (the top panel of Table 2) are the same because the estimations of the individually calculated AUC approach and the proposed mixed model are not affected by the value of β_1^G. The rest of Table 3 has similar patterns as Table 2, showing that when there are missing data, the proposed method outperforms the two-step methods under the simulation scenarios with balanced baseline between the two treatment groups.

We provide power curves for different sample sizes ($n = 100$ to 400) under different missing mechanism and missing patterns, using the simulated data with a linear decline trend and unbalanced baseline in Appendix B. It shows that the proposed method has bigger power than the comparison methods in all scenarios.

5 | DATA APPLICATION

We first applied the mixed model for the two biomarkers of the CENIC study, 3-HPMA and PheT without transformation with a model similar to Equation (1) except that the outcome variable was Y_{ij}. The residual Q-Q plots (left panel of Figure A2 in Appendix A) showed that the normality assumption for the error term was not plausible. We then applied the proposed mixed effects model on the log-transformed biomarker in Equation (1). The residual Q-Q plots of the proposed model (right panel of Figure A1) and the residual plot (right panel of Figure A2) confirmed that the log-transformation was appropriate for the two biomarkers. The mixed model with the log-transformed biomarker then takes the same form as Equation (1) but with $m = 5$, $k = 2$:

$$\log(Y_{i0}) = \beta_0 + \beta_1^G I(G_i = 1) + \beta_2^G I(G_i = 2) + b_i + e_{i0},$$
$$\log(Y_{i1}) = \beta_0 + \beta_1^J + \beta_1^G I(G_i = 1) + \beta_2^G I(G_i = 2) + \beta_{11}^{GJ} I(G_i = 1) + \beta_{21}^{GJ} I(G_i = 2) + b_i + e_{i1},$$
$$\cdots$$
$$\log(Y_{i5}) = \beta_0 + \beta_5^J + \beta_1^G I(G_i = 1) + \beta_2^G I(G_i = 2) + \beta_{15}^{GJ} I(G_i = 1) + \beta_{25}^{GJ} I(G_i = 2) + b_i + e_{i5}.$$

Two comparison models, the LOCF and the MI methods, using individually imputed or calculated AUC's in the outcome variable, were also applied to the data. For the MI method, we followed the published CENIC study[5] to include a number of baseline demographic variables and smoking variables as auxiliary variables and applied two different approaches: (a) impute the components of the AUC, that is, the biomarker values at individual visits,

with the MCMC method, then calculate the AUC (this method was used in the simulation study and also as the primary approach in the published study); (b) impute the composite outcome AUC together with its components, with a fully conditional specification (FCS) method, where linear regression was used for continuous variables and logistic regression for categorical variables. As shown in Table 4, the results of these two approaches were similar.

To adjust for the potential baseline imbalance in the biomarkers when using the LOCF or MI method, we used two ad hoc approaches. In the first approach, we followed the simulation studies to use $\log\{AUC_i/(t_m Y_{i0})\}$ as the outcome variable in linear regression (referred to as Linear Regression Model 1 or LRM1). In the second approach (also the primary approach in the published CENIC study[5]), we adjusted the log-transformed baseline level of the corresponding biomarker as a covariate (referred to as Linear Regression Model 2 or LRM2). The analysis result using the proposed mixed model (the SAS macro for this model is available in Data S1) and the linear regression models with individually imputed and calculated AUC's are shown in Table 4.

The results in Table 4 show that the MI method and the proposed mixed model provided similar point estimates of the AUC ratio between the immediate and gradual reduction groups, while the LOCF method gave a slightly underestimated treatment effect (toward null, ie, ratio = 1). Based on the mixed model, there was a 20% (95% CI, 15-26%) and 15% (95% CI, 10-19%) reduction in the 20-week cumulative exposure of 3-HPMA and PheT, respectively, comparing the immediate and gradual reduction groups. Similar amounts of reduction in 3-HPMA (18%) and PheT (14%) were found between the immediate reduction and control groups based on all three methods. No significant difference was found between the gradual reduction and the control groups. We also found that, compared with the LOCF method, the mixed model and the MI method provided larger estimated treatment effects (ie, the estimated AUC ratios were further from 1.0) of the immediate reduction group vs the gradual reduction or the control group. Between the mixed model and the MI method, the former provided a slightly narrower 95% CI. These findings are consistent with the simulation results.

6 | DISCUSSION

Missing data is a frequently encountered problem in longitudinal studies. The estimation for AUC is difficult when there are missing data in repeatedly measured response variables. A simple two-step approach is to first impute the missing values, using simple or multiple imputation methods, and then calculate the AUC for each individual based on the imputed data. The individually calculated AUCs can then be used as response variables to make group comparisons. In this paper, we proposed a mixed effects model approach, which circumvented the calculation of individual AUCs by fitting a mixed effects model on repeatedly measured, log-normally distributed biomarker data from all individuals, where the between-group difference in log-transformed AUC can be expressed as a function of regression parameters from the mixed model. The proposed mixed effects model showed negligible bias for various studied scenarios and were shown to provide larger powers and narrower 95% CIs than the LOCF and MI methods. The proposed method was also shown

to be flexible enough to handle data with different missing patterns or no missing, with or without baseline imbalance. In addition, the computing of the proposed method was found more efficient and easier to implement than the two-step approach, especially with the computationally intensive multiple imputation method.

We recognize that there exists the situation where the response variable has a skewed distribution while the covariate of interest has an absolute effect rather than a relative effect on the response variable. In this case, a linear form of the model rather than a log-linear form, as the one studied in this paper, would be more appropriate. A maximum likelihood (ML)-based approach has been studied for the univariate case,[26] which may be extended to the repeated measure case and applied to draw inference on the absolute effect of the covariate on the AUC of the response variable. It will be interesting to investigate the relative performance of such a model compared with the existing linear mixed model.[12]

Note that the proposed method for estimating the treatment effects in the AUC is built on a linear mixed model (for log-transformed biomarkers), which is known for its robustness in terms of the mean model estimation with respect to misspecification of the random effects distribution.[27] However, a thorough investigation of the robustness of the proposed method for the estimation of the treatment effects on the AUC can be a future research direction. We also note that a GEE marginal model is a robust and popular competitor to the linear mixed model and that when the outcome is continuous, both methods' regression coefficients in the mean model have a marginal interpretation. However, the proposed mixed effects model based approach has the advantage that the inference of the baseline-scaled AUC can be easily obtained by using the SAS NLMIXED procedure (see the SAS macro in Data S1), which we think is an appealing feature in applications.

Our proposed model assumed that subjects in the study had the same sequence of visit times and the outcome variables were log-normally distributed. Future studies can extend the model to deal with different visit-time sequences and biomarkers with other distributions. Additionally, as the MI and many other missing data methods, the proposed mixed effects model requires the data to be MAR, which includes the MCAR as a special case. However, the missing mechanism could be MNAR in applications, especially for the dropout, and it is well known that the observed data alone does not allow us to distinguish between MAR and MNAR. A potential way to deal with MNAR mechanism is to use joint models for longitudinal and time-to-event data by explicitly modeling the dropout process in a survival sub-model.[28,29] A future research direction could be using the joint modeling approach to make inference on the AUCs, while taking into account the MNAR data. Finally, in our data there were no biomarker values below a detection limit, hence log-transformation was a valid method to use. However, it is not uncommon to observe biomarkers below a detection limit, referred to as left-censoring in statistical literature, in which case, the computation of the AUC can be complicated. A likelihood approach has been proposed to explicitly account for the left-censoring of markers in a linear mixed model based on the nontransformed response variable.[30] Extension of such a model for a log-transformed response variable is warranted.

DATA AVAILABILITY STATEMENT

The data are available to investigators with a formal request to Dorothy K. Hatsukami, the principle investigator of the parent trial.[5]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Appendix

APPENDIX A.

Model diagnosis plots for the cenic data

The quantile-quantile (Q-Q) plots of the two biomarkers (3-HPMA and PheT) and their log-transformed values at baseline are shown in Figure A1. The residual Q-Q plots of the linear mixed model for the original scale of the biomarkers by using Bell et al. model[12] and for the log-transformed biomarkers by using the proposed model are shown in Figure A2.
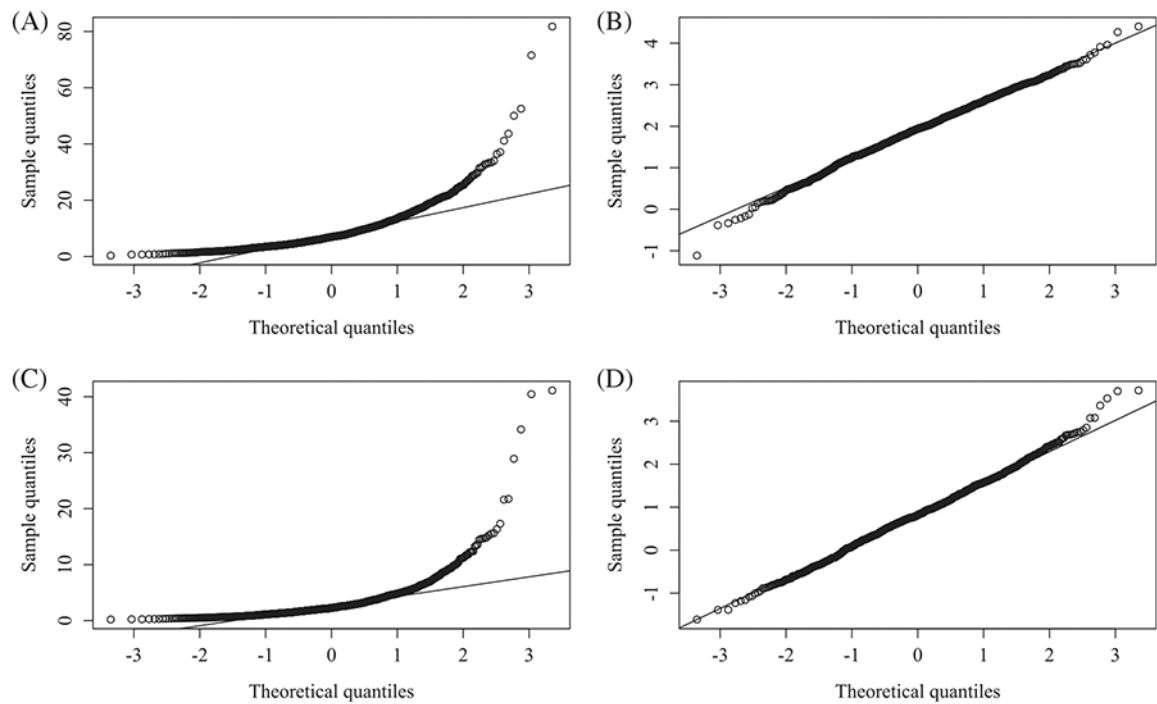
FIGURE A1.

Normal Q-Q plots for biomarkers and log-transformed biomarkers at baseline for the CENIC study: A, 3-HPMA; B, log (3-HPMA); C, PheT; D, log(PheT)
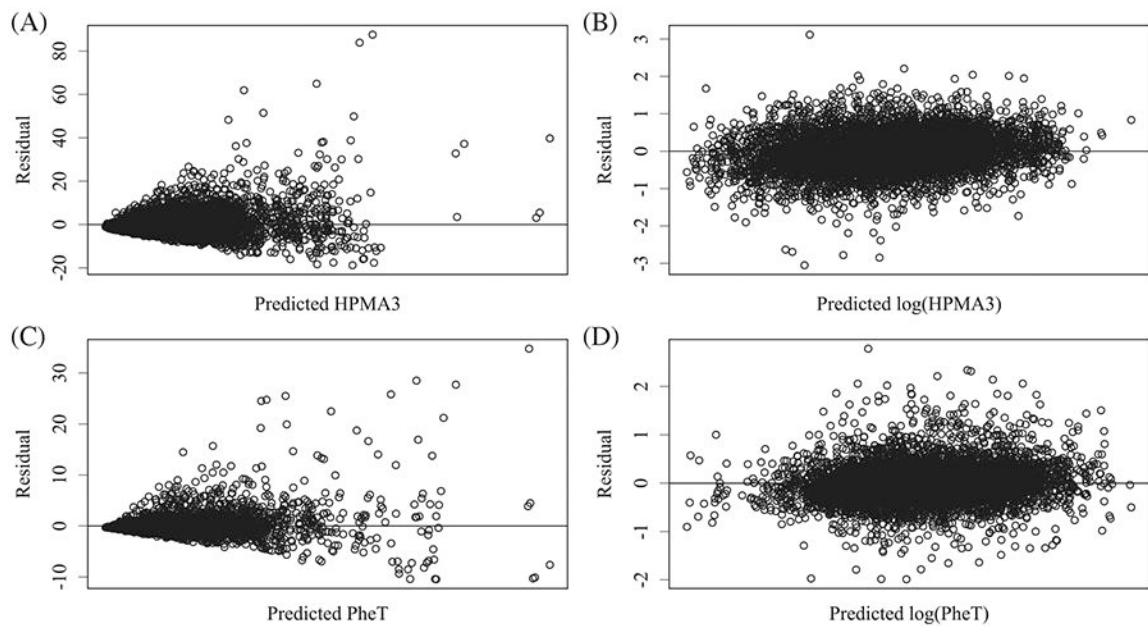
FIGURE A2.

Residual plots for the mixed models with biomarkers and log-transformed biomarkers in the CENIC study as the outcome variable: A, 3-HPMA; B, log(3-HPMA); C, PheT; D, log(PheT)

APPENDIX B.

Power curve for simulations

We present the power curves for different missing mechanism and missing patterns, using the simulated data with a linear decline trend and unbalanced baseline. The power curves for the simulated data with no missing are presented in Figure B1, and those for MCAR and MAR are presented in Figures B2 and B3, respectively.
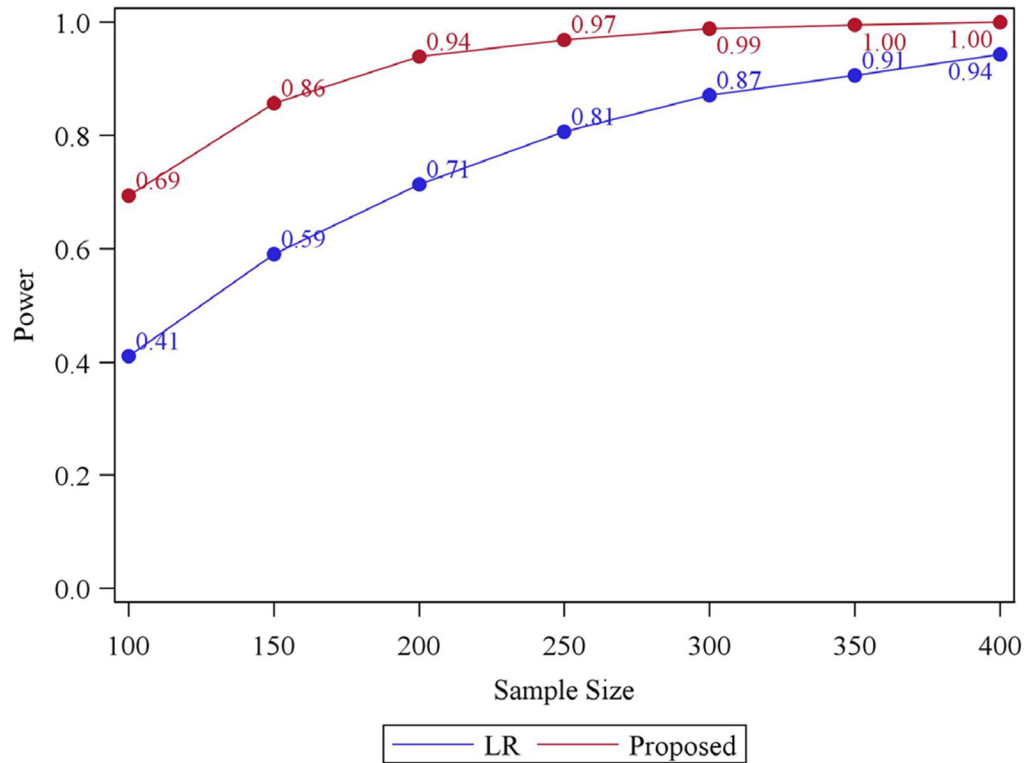
FIGURE B1.
Power curve for simulation scenarios with no missing data, linear decline trend with unbalanced baseline; LR, linear regression; Proposed, the proposed method

FIGURE B2.
Power curve for simulation scenarios under MCAR, linear decline trend with unbalanced baseline: A, Low missing rate, dropout missing; B, Low missing rate, intermittent missing; C, High missing rate, dropout missing; D, High missing rate, intermittent missing; LOCF, last observation carried forward; MI, multiple imputation; Proposed, the proposed method
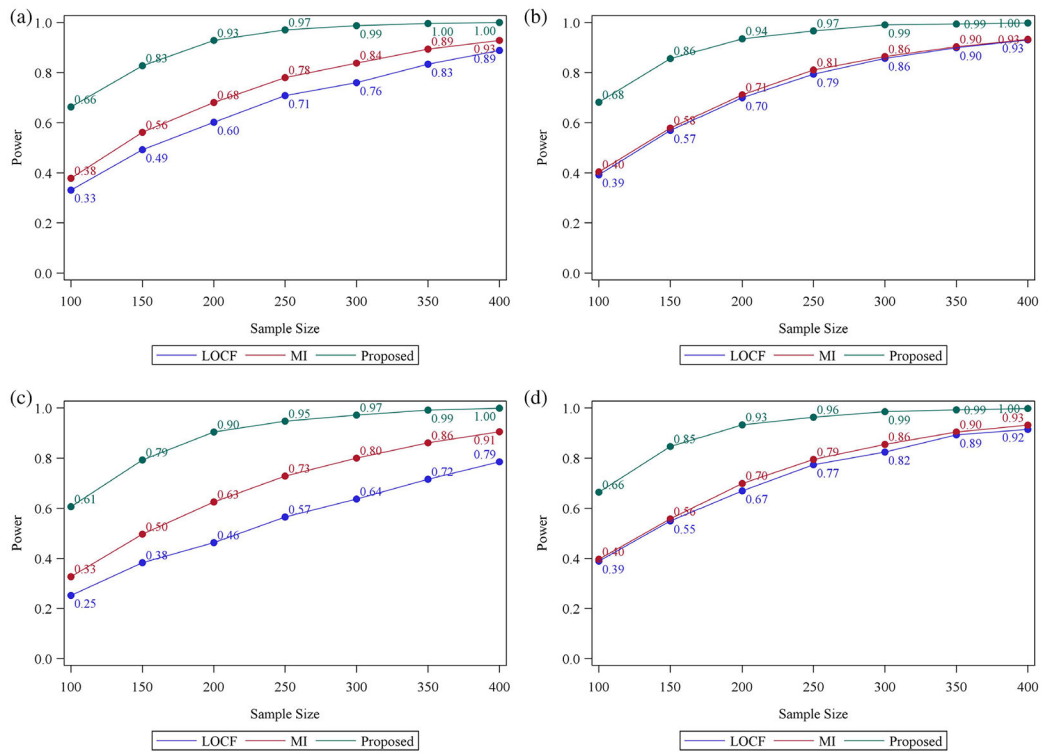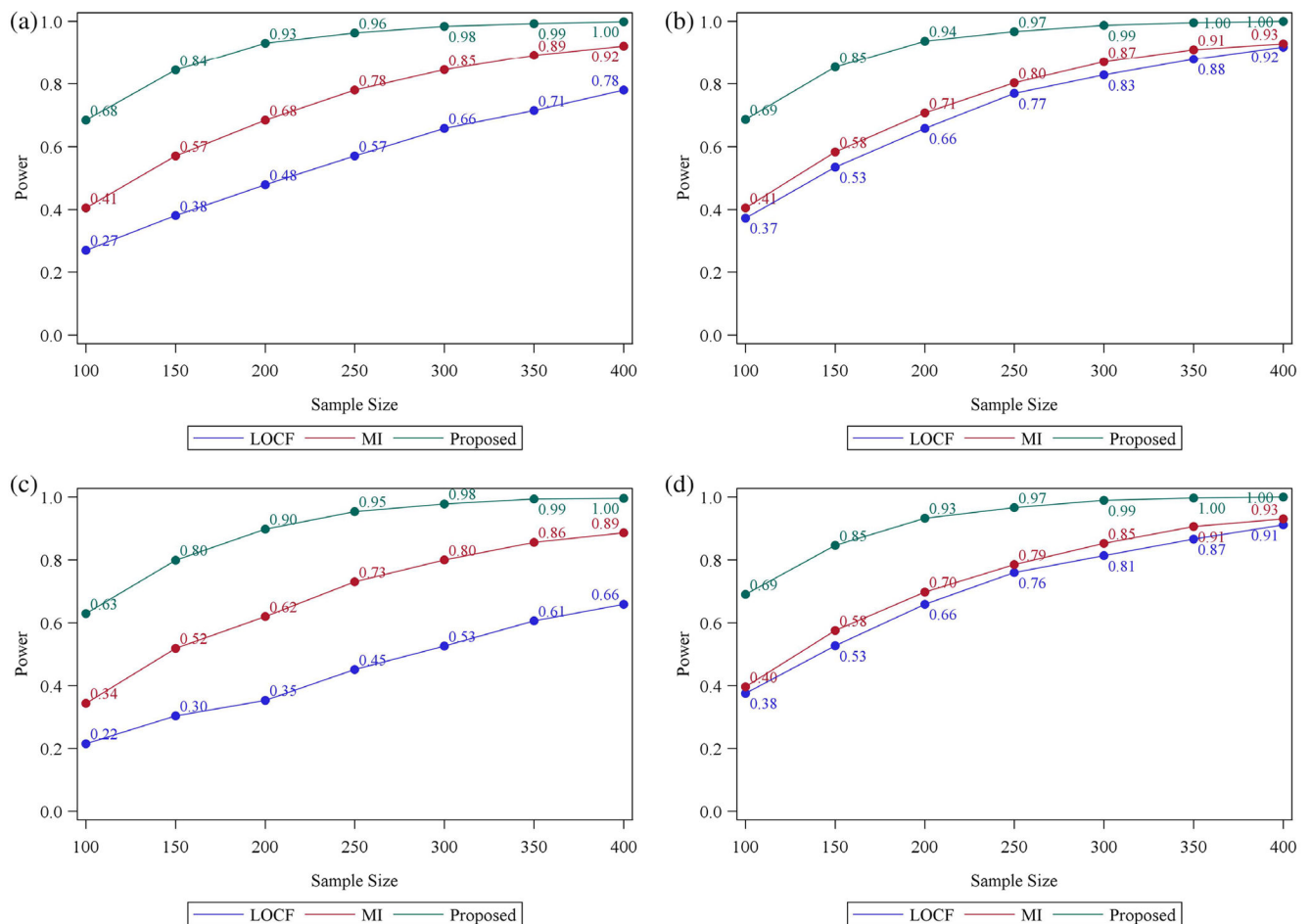
FIGURE B3.
Power curve for simulation scenarios under MAR, linear decline trend with unbalanced baseline: A, Low missing rate, dropout missing; B, Low missing rate, intermittent missing; C, High missing rate, dropout missing; D, High missing rate, intermittent missing; LOCF, last observation carried forward; MI, multiple imputation; Proposed, the proposed method

REFERENCES

1. Hatsukami DK, Zaatari G, Donny E. The case for the WHO advisory note, global nicotine reduction strategy. Tob Control. 2017;26: e29–e30. [PubMed: 27357938]

2. Benowitz NL, Henningfield JE. Reducing the nicotine content to make cigarettes less addictive. Tob Control. 2013;22:i14–i17. [PubMed: 23591498]

3. Hatsukami DK, Donny EC, Koopmeiners JS, Benowitz NL. Compensatory smoking from gradual and immediate reduction in cigarette nicotine content. Cancer Epidemiol Biomarkers Prev. 2015;24:472–476. [PubMed: 25515551]

4. Hatsukami DK, Kotlyar M, Hertsgaard LA, et al. Reduced nicotine content cigarettes: effects on toxicant exposure, dependence and cessation. Addiction. 2010;105:343–355. [PubMed: 20078491]

5. Hatsukami DK, Luo X, Jensen JA, et al. Effect of immediate vs gradual reduction in nicotine content of cigarettes on biomarkers of smoke exposure: a randomized clinical trial. Jama. 2018;320:880–891. [PubMed: 30193275]

6. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2nd ed. Hoboken, NJ: Wiley; 2011.

7. Wilding GE, Chandrasekhar R, Hutson AD. A new linear model-based approach for inferences about the mean area under the curve. Stat Med. 2012;31:3563–3578. [PubMed: 23175104]

8. Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology. 2003;28:916–931. [PubMed: 12892658]

9. Le J, Bradley JS, Murray W, et al. Improved vancomycin dosing in children using area under the curve exposure. Pediatr Infect Dis J. 2013;32:e155–e163. [PubMed: 23340565]

10. Yeh KC, Kwan KC. A comparison of numerical integrating algorithms by trapezoidal, lagrange, and spline approximation. J Pharmacokinet Biopharm. 1978;6:79–98. [PubMed: 650423]

11. Daniels MJ, Hogan JW. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Boca Raton, FL: Chapman Hall/CRC; 2008.

12. Bell ML, King MT, Fairclough DL. Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data: summary measures versus summary statistics. SAGE Open. 2014;4(2):1–12. 10.1177/2158244014534858

13. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd ed. Hoboken, NJ: Wiley; 2002.

14. Spritzler J, DeGruttola VG, Pei L. Two-sample tests of area-under-the-curve in the presence of missing data. Int J Biostat. 2008;4(1):1. 10.2202/1557-4679.1068

15. Gustavsson S, Fagerberg B, Sallsten G, Andersson E. Regression models for log-normal data: comparing different methods for quantifying the association between abdominal adiposity and biomarkers of inflammation and insulin resistance. Int J Environ Res Public Health. 2014;11:3521–3539. [PubMed: 24681553]

16. Grund B, Sabin C. Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. Curr Opin HIV AIDS. 2006; 5:473–479.

17. Cheng YC, Reyes-Guzman C, Christensen C, et al. Biomarkers of exposure among adult smokeless tobacco users in the population assessment of tobacco and health study (wave 1, 2013–2014). Curr Opin HIV AIDS. 2020;29:659–667.

18. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. Bioscience. 2001;51:341–352.

19. Halvorsen R, Palmquist R. The interpretation of dummy variables in semilogarithmic equations. Am Econ Rev. 1980;70:474–475.

20. Food and Drug Administration (FDA). E9 Statistical Principles for Clinical Trials. 9 16, 1998. https://www.hhs.gov/guidance/document/e9-statistical-principles-clinical-trials

21. Food and Drug Administration (FDA), Center for Drug Evaluation and Research. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes Guidance for Industry. 4, 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinicaltrials-drugs-and-biologics-continuous-outcomes-guidance

22. Senn S, Stevens L, Chaturvedi N. Repeated measures in clinical trials: simple strategies for analysis using summary measures. Stat Med. 2000;19:861–877. [PubMed: 10734289]

23. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press; 2010.

24. Siddiqui O, Ali M. A comparison of the random-effects pattern mixture model with last-observation-carried-forward (LOCF) analysis in longitudinal clinical trials with dropouts. J Biopharm Stat. 1998;8:545–563. [PubMed: 9855033]

25. Schafer JL. Analysis of Incomplete Multivariate Data. New York, NY: Chapman Hall/CRC; 1997.

26. Gustavsson S, Johannesson S, Sallsten G, Andersson E. Linear maximum likelihood regression analysis for untransformed log-normally distributed data. Open J Stat. 2012;2:389–400.

27. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York, NY: Springer; 2000.

28. Vonesh E, Greene T, Schluchter M. Shared parameter models for the joint analysis of longitudinal data and event times. Stat Med. 2006; 25:143–163. [PubMed: 16025541]

29. Rizopoulos D Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. Boca Raton, FL: Chapman Hall/CRC; 2012.

30. Chandrasekhar R, Shi Y, Hutson A, Wilding G. Likelihood-based inferences about the mean area under a longitudinal curve in the presence of observations subject to limits of detection. Pharm Stat. 2015;14:252–261. [PubMed: 25832442]
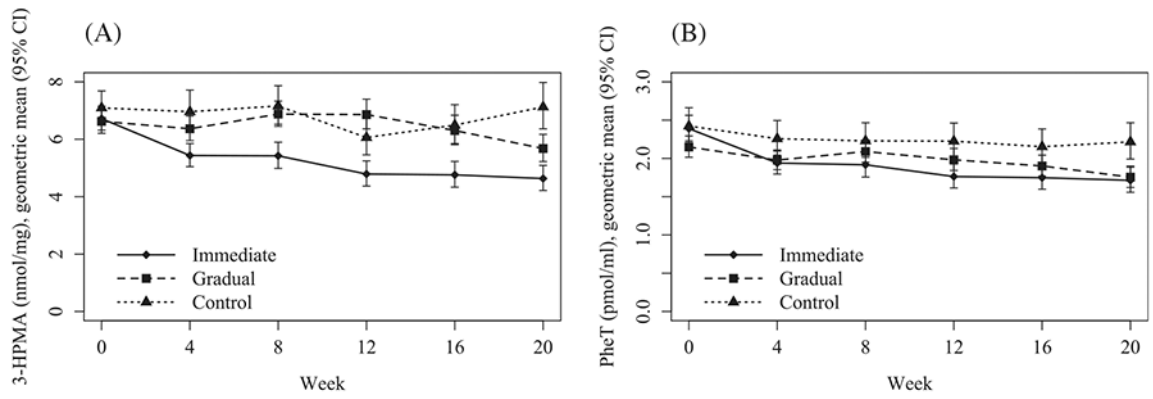
FIGURE 1.
Geometric mean and 95% confidence interval (CI) of biomarker level at each visit by treatment group for the CENIC study: A, 3-HPMA; B, PheT. The line chart with median and boxplot with mean, interquartile range, and other statistics of the same data were presented by Hatsukami et al.[5]
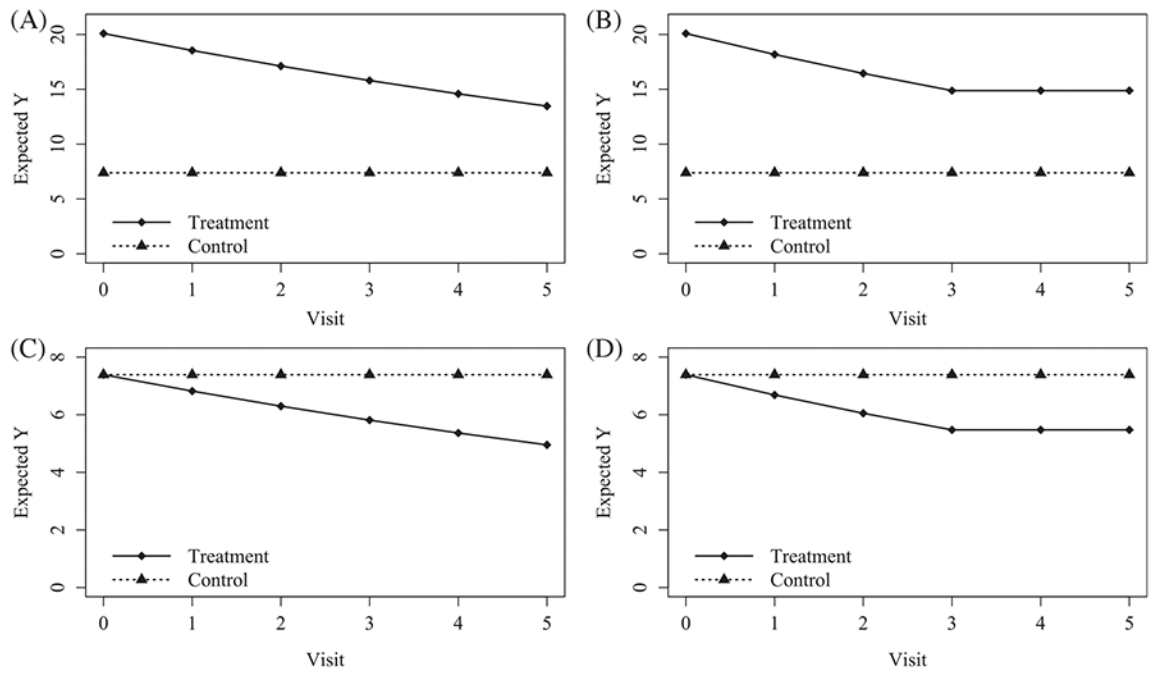
FIGURE 2.
Trajectory of E(*Y*) by group for simulated data with, A, a linear decline trend (in E[log(*Y*ᵢⱼ)]) for treatment with unbalanced baseline; B, a linear spline decline for treatment with unbalanced baseline; C, a linear decline for treatment with balanced baseline; D, a linear spline decline for treatment with balanced baseline

TABLE 1

Summary of missing data of 3-HPMA and PheT of the CENIC study

Week	3-HPMA				PheT			
	Overall *n* (%)	Immediate *n* (%)	Gradual *n* (%)	Control *n* (%)	Overall *n* (%)	Immediate *n* (%)	Gradual *n* (%)	Control *n* (%)
0	0	0	0	0	4 (<1)	1 (<1)	2 (<1)	1 (<1)
4	133 (11)	86 (17)	31 (6)	16 (6)	134 (11)	86 (17)	32 (6)	16 (6)
8	210 (17)	124 (25)	57 (11)	29 (12)	201 (16)	122 (24)	53 (11)	26 (10)
12	261 (21)	147 (29)	80 (16)	34 (14)	249 (20)	143 (28)	75 (15)	31 (12)
16	293 (23)	162 (32)	92 (18)	39 (16)	279 (22)	155 (31)	88 (18)	36 (14)
20	312 (25)	171 (34)	102 (20)	39 (16)	295 (24)	161 (32)	95 (19)	39 (16)
Missing pattern								
Baseline	0	0	0	0	4 (<1)	1 (<1)	2 (<1)	1 (<1)
Intermittent	53 (4)	22 (4)	19 (4)	12 (5)	22 (2)	9 (2)	9 (2)	4 (2)
Dropout	312 (25)	171 (34)	102 (20)	39 (16)	295 (24)	161 (32)	95 (19)	39 (16)

Abbreviations: Baseline, missing at the baseline visit; Control, conventional level of nicotine group; Dropout, dropout missing; Gradual, gradual nicotine reduction group; Immediate, immediate nicotine reduction group; Intermittent, intermittent missing.

TABLE 2

Simulation result for the comparison of the LOCF, MI, and mixed model for data with unbalanced baseline between groups

Missing pattern	Missing mech [a]	Method [b]	Linear decline trend					Linear spline decline trend				
			Coverage (%)	Power	Bias (%) [c]	Empirical SD [d]	Mean SE [e]	Coverage (%)	Power	Bias (%) [c]	Empirical SD [d]	Mean SE [e]
No missing												
—	—	LR	94.0	0.71	−1.13	0.076	0.076	93.8	0.76	−1.23	0.076	0.076
		Proposed	94.5	0.94	−1.10	0.055	0.053	94.4	0.96	−1.02	0.054	0.053
Low missing rate												
Dropout	MCAR	LOCF	94.1	0.60	−12.42	0.076	0.076	94.2	0.67	−10.76	0.075	0.076
		MI	94.6	0.68	−1.01	0.079	0.080	94.6	0.73	−1.14	0.079	0.080
		Proposed	94.7	0.93	−1.15	0.056	0.055	94.7	0.95	−1.08	0.056	0.055
	MAR	LOCF	89.9	0.48	−24.43	0.079	0.077	90.0	0.55	−21.75	0.079	0.077
		MI	93.6	0.68	−0.51	0.079	0.080	93.9	0.74	−0.65	0.079	0.080
		Proposed	94.5	0.93	−1.22	0.056	0.054	94.4	0.96	−1.18	0.056	0.054
Intermittent	MCAR	LOCF	93.8	0.70	−3.43	0.076	0.076	93.7	0.75	−3.38	0.076	0.076
		MI	93.9	0.71	−1.06	0.077	0.077	93.9	0.76	−1.17	0.077	0.077
		Proposed	94.7	0.94	−0.90	0.055	0.053	94.9	0.96	−0.84	0.055	0.053
	MAR	LOCF	94.0	0.66	−6.98	0.076	0.076	93.8	0.71	−6.58	0.076	0.076
		MI	94.0	0.71	−1.15	0.077	0.077	93.8	0.78	−11.23	0.077	0.078
		Proposed	94.6	0.94	−1.00	0.055	0.053	94.6	0.96	−0.91	0.055	0.053
High missing rate												
Dropout	MCAR	LOCF	89.1	0.46	−26.22	0.076	0.076	89.8	0.55	−22.93	0.076	0.076
		MI	95.0	0.63	−0.68	0.085	0.085	94.8	0.67	−0.86	0.085	0.085
		Proposed	94.5	0.90	−1.27	0.059	0.058	94.5	0.93	−1.19	0.059	0.057
	MAR	LOCF	84.4	0.35	−36.70	0.078	0.076	85.9	0.44	−32.15	0.078	0.076
		MI	94.6	0.62	−1.38	0.085	0.085	94.6	0.66	−1.52	0.085	0.085
		Proposed	94.7	0.90	−1.70	0.059	0.057	94.6	0.93	−1.55	0.058	0.056
Intermittent	MCAR	LOCF	94.4	0.67	−6.77	0.076	0.075	94.3	0.73	−6.59	0.076	0.075
		MI	94.1	0.70	−0.84	0.078	0.078	94.0	0.74	−0.96	0.078	0.078
		Proposed	95.0	0.93	−0.86	0.056	0.054	95.0	0.95	−0.80	0.055	0.054
	MAR	LOCF	93.7	0.66	−8.93	0.076	0.076	93.5	0.71	−8.29	0.075	0.076

Missing pattern	Missing mech[a]	Method[b]	Linear decline trend					Linear spline decline trend				
			Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]	Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]
		MI	93.6	0.70	−0.95	0.078	0.078	93.9	0.75	−1.07	0.078	0.078
		Proposed	94.0	0.93	−1.19	0.055	0.054	94.3	0.96	−1.09	0.055	0.054

[a]Missing mechanism: missing completely at random (MCAR) and missing at random (MAR).

[b]LR: linear regression with individually calculated AUC (with no imputation because there are no missing data) used in the transformation of the form $\log\{AUC_i/(t_m Y_i0)\}$ as the outcome variable; Proposed: the mixed effects model for data with no missing or missing under different missing mechanisms and patterns; LOCF and MI: the two-step approach with the missing data imputed by the last observation carried forward simple imputation and multiple imputation methods, then the imputed data were analyzed using the LR method.

[c]Relative bias calculated as 100 × {mean estimated log(baseline-scaled AUC ratio)−log(true ratio)}/log(true ratio).

[d]Monte-Carlo SD of the estimated log(baseline-scaled AUC ratio).

[e]Mean of the SE of the estimated log(baseline-scaled AUC ratio).

TABLE 3

Simulation result for the comparison of the LOCF, MI, and mixed model for data with balanced baseline between groups

Missing pattern	Missing mech[a]	Method[b]	Linear decline trend					Linear spline decline trend				
			Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]	Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]
No missing												
—	—	LR	94.0	0.71	−1.13	0.076	0.076	93.8	0.76	−1.23	0.076	0.076
		Proposed	94.5	0.94	−1.10	0.055	0.053	94.4	0.96	−1.02	0.054	0.053
Low missing rate												
Dropout	MCAR	LOCF	94.1	0.60	−12.42	0.076	0.076	94.2	0.67	−10.76	0.075	0.076
		MI	94.6	0.68	−1.01	0.079	0.080	94.6	0.73	−1.14	0.079	0.080
		Proposed	94.7	0.93	−1.15	0.056	0.055	94.7	0.95	−1.08	0.056	0.055
	MAR	LOCF	93.3	0.61	−11.87	0.077	0.077	93.6	0.68	−9.36	0.076	0.077
		MI	93.6	0.67	−0.66	0.080	0.081	93.7	0.73	−0.73	0.080	0.081
		Proposed	94.6	0.93	−0.93	0.056	0.055	94.5	0.95	−0.76	0.056	0.055
Intermittent	MCAR	LOCF	93.8	0.70	−3.43	0.076	0.076	93.7	0.75	−3.38	0.076	0.076
		MI	93.9	0.71	−1.06	0.077	0.077	93.9	0.76	−1.17	0.077	0.077
		Proposed	94.7	0.94	−0.90	0.055	0.053	94.9	0.96	−0.84	0.055	0.053
	MAR	LOCF	94.1	0.72	−2.00	0.076	0.076	94.1	0.77	−1.85	0.075	0.076
		MI	93.7	0.72	−1.22	0.077	0.077	93.6	0.76	−1.36	0.077	0.077
		Proposed	94.8	0.94	−1.16	0.055	0.053	94.7	0.96	−1.10	0.054	0.053
High missing rate												
Dropout	MCAR	LOCF	89.1	0.46	−26.22	0.076	0.076	89.8	0.55	−22.93	0.076	0.076
		MI	95.0	0.63	−0.68	0.085	0.085	94.8	0.67	−0.86	0.085	0.085
		Proposed	94.5	0.90	−1.27	0.059	0.058	94.5	0.93	−1.19	0.059	0.057
	MAR	LOCF	91.1	0.50	−23.90	0.076	0.076	91.9	0.59	−20.01	0.076	0.076
		MI	94.3	0.61	−1.37	0.085	0.086	94.4	0.66	−1.72	0.085	0.086
		Proposed	94.3	0.91	−1.36	0.059	0.057	94.3	0.93	−1.37	0.059	0.057
Intermittent	MCAR	LOCF	94.4	0.67	−6.77	0.076	0.075	94.3	0.73	−6.59	0.076	0.075
		MI	94.1	0.70	−0.84	0.078	0.078	94.0	0.74	−0.96	0.078	0.078
		Proposed	95.0	0.93	−0.86	0.056	0.054	95.0	0.95	−0.80	0.055	0.054
	MAR	LOCF	94.5	0.70	−4.35	0.076	0.076	94.7	0.76	−3.93	0.076	0.075

Missing pattern	Missing mech[a]	Method[b]	Linear decline trend					Linear spline decline trend				
			Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]	Coverage (%)	Power	Bias (%)[c]	Empirical SD[d]	Mean SE[e]
		MI	94.5	0.70	−0.73	0.078	0.078	94.3	0.75	−0.81	0.078	0.078
		Proposed	94.0	0.93	−1.05	0.055	0.054	93.9	0.95	−1.10	0.055	0.054

[a] Missing mechanism: missing completely at random (MCAR) and missing at random (MAR).

[b] LR: linear regression with individually calculated AUC (with no imputation because there are no missing data) used in the transformation of the form log{ $AUC_i/(t_mY_{i0})$ } as the outcome variable; Proposed: the mixed effects model for data with no missing or missing under different missing mechanisms and patterns; LOCF and MI: the two-step approach with the missing data imputed by the last observation carried forward simple imputation and multiple imputation methods, then the imputed data were analyzed using the LR method.

[c] Relative bias calculated as 100 × {mean estimated log(baseline-scaled AUC ratio) −log(true ratio) }/log(true ratio).

[d] Monte-Carlo SD of the estimated log(baseline-scaled AUC ratio).

[e] Mean of the SE of the estimated log(baseline-scaled AUC ratio).

TABLE 4

Analysis of area under the curve for the CENIC study

Model[a]	Biomarker	Immediate vs gradual		Immediate vs control		Gradual vs control	
		Est. ratio (95% CI)[b]	P	Est. ratio (95% CI)	P	Est. ratio (95% CI)	P
LOCF	3-HPMA	0.83 (0.78, 0.89)	<0.001	0.83 (0.76, 0.90)	<0.001	0.99 (0.92, 1.08)	0.89
(LRM1)	PheT	0.88 (0.84, 0.93)	<0.001	0.89 (0.84, 0.96)	<0.001	1.01 (0.95, 1.08)	0.71
LOCF	3-HPMA	0.84 (0.79, 0.89)	<0.001	0.81 (0.76, 0.87)	<0.001	0.97 (0.90, 1.04)	0.35
(LRM2)	PheT	0.90 (0.86, 0.95)	<0.001	0.89 (0.84, 0.95)	<0.001	0.99 (0.93, 1.05)	0.65
MI (cpnt)	3-HPMA	0.82 (0.76, 0.89)	<0.001	0.83 (0.76, 0.91)	<0.001	1.01 (0.93, 1.11)	0.77
(LRM1)	PheT	0.86 (0.81, 0.91)	<0.001	0.87 (0.81, 0.93)	<0.001	1.01 (0.94, 1.08)	0.79
MI (cpnt)	3-HPMA	0.83 (0.77, 0.88)	<0.001	0.81 (0.75, 0.88)	<0.001	0.98 (0.91, 1.06)	0.64
(LRM2)	PheT	0.88 (0.83, 0.93)	<0.001	0.86 (0.81, 0.92)	<0.001	0.98 (0.92, 1.04)	0.52
MI (comp)	3-HPMA	0.82 (0.76, 0.88)	<0.001	0.82 (0.75, 0.90)	<0.001	1.01 (0.92, 1.10)	0.87
(LRM1)	PheT	0.86 (0.81, 0.91)	<0.001	0.87 (0.80, 0.94)	<0.001	1.01 (0.94, 1.09)	0.78
MI (comp)	3-HPMA	0.82 (0.77, 0.87)	<0.001	0.80 (0.75, 0.86)	<0.001	0.98 (0.91, 1.05)	0.52
(LRM2)	PheT	0.88 (0.83, 0.93)	<0.001	0.86 (0.80, 0.93)	<0.001	0.98 (0.92, 1.05)	0.54
Proposed	3-HPMA	0.80 (0.74, 0.85)	<0.001	0.81 (0.74, 0.89)	<0.001	1.02 (0.94, 1.11)	0.64
	PheT	0.85 (0.81, 0.90)	<0.001	0.86 (0.80, 0.92)	<0.001	1.00 (0.94, 1.07)	0.93

[a]LOCF and MI: the two-step approach with the missing data imputed by the last observation carried forward (LOCF) simple imputation or the multiple imputation (MI) method; MI (cpnt): impute the components of AUC with MCMC method then calculate the AUC; MI (comp): impute the composite outcome AUC together with its components, with the FCS method; LRM1: linear regression model 1, where the outcome variable was $\log\{AUC_i/(t_m Y_{i0})\}$; LRM2: linear regression model 2, where baseline log-biomarker was adjusted as a covariate; Proposed: the proposed mixed effects model.

[b]The estimated ratio of AUC (baseline-scaled AUC for the proposed model) between two treatment groups and 95% confidence interval (CI).