



OPEN

Longitudinal hierarchical Bayesian models of covariate effects on airway and alveolar nitric oxide

Jingying Weng, Noa Molshatzki, Paul Marjoram, W. James Gauderman, Frank D. Gilliland & Sandrah P. Eckel

Biomarkers such as exhaled nitric oxide (FeNO), a marker of airway inflammation, have applications in the study of chronic respiratory disease where longitudinal studies of within-participant changes in the biomarker are particularly relevant. A cutting-edge approach to assessing FeNO, called multiple flow FeNO, repeatedly assesses FeNO across a range of expiratory flow rates at a single visit and combines these data with a deterministic model of lower respiratory tract NO to estimate parameters quantifying airway wall and alveolar NO sources. Previous methodological work for multiple flow FeNO has focused on methods for data from a single participant or from cross-sectional studies. Performance of existing ad hoc two-stage methods for longitudinal multiple flow FeNO in cohort or panel studies has not been evaluated. In this paper, we present a novel longitudinal extension to a unified hierarchical Bayesian (L_U_HB) model relating longitudinally assessed multiple flow FeNO to covariates. In several simulation study scenarios, we compare the L_U_HB method to other unified and two-stage frequentist methods. In general, L_U_HB produced unbiased estimates, had good power, and its performance was not sensitive to the magnitude of the association with a covariate and correlations between NO parameters. In an application relating height to longitudinal multiple flow FeNO in schoolchildren without asthma, unified analysis methods estimated positive, statistically significant associations of height with airway and alveolar NO concentrations and negative associations with airway wall diffusivity while estimates from two-stage methods were smaller in magnitude and sometimes non-significant.

Sophisticated statistical methods are needed to link longitudinal assessments of a biomarker to patient-level characteristics and time-varying exposures in the context of a deterministic mathematical model describing the production and dynamics of the biomarker within the human body. This paper presents statistical methods developed for such longitudinal assessments of the fractional concentration of exhaled nitric oxide, FeNO, a biomarker of airway inflammation used in clinical^{1–3} and epidemiological research^{4–6}.

FeNO is an exhaled breath biomarker conventionally assessed at the target expiratory flow rate of 50 ml/s (FeNO₅₀)⁷. A cutting-edge approach, called multiple flow FeNO, repeatedly assesses FeNO across a range of expiratory flow rates and combines these data with a deterministic model of NO in the lower respiratory tract to estimate parameters quantifying the effect of airway wall and alveolar sources. Literature on the modeling of multiple flow FeNO data has focused on methods for data collected from one person at a single visit^{8–11}. Multiple flow FeNO data present statistical challenges which require sophisticated statistical methods. Many studies of multiple flow FeNO conduct analyses using a two-stage approach: (1) estimate airway and alveolar NO parameters and (2) treat the estimated NO parameters as observed outcomes in linear regressions relating NO parameters to factors of interest (asthma medication use, air pollution exposures, etc.). In a previous paper, we presented a novel unified hierarchical Bayesian (U_HB) model for estimating *cross-sectional* associations of covariates with NO parameters using data from a single multiple flow FeNO test session for each study participant. We also found, in an extensive simulation study, that the U_HB method was less biased and had better power/type I error compared to conventional two-stage methods¹².

There is a need for longitudinal data analysis methods for FeNO. Biomarkers like FeNO are particularly promising for tracking within-person changes over time since they tend to be relatively stable within persons, despite considerable heterogeneity across people^{13–16}. Longitudinal trends in study populations with repeated measures of the conventional FeNO₅₀ are generally modeled using standard longitudinal data analysis techniques,

Department of Population and Public Health Sciences, University of Southern California, 2001 N. Soto Street, SSB 202B, MC-9234, Los Angeles, CA 90089, USA. email: eckel@usc.edu

such as linear mixed effects models (LMM) or generalized additive linear mixed effects models (GAMM)¹⁷. Panel studies or longitudinal cohort studies with repeated measures of multiple flow FeNO across multiple visits can highlight within-person trends in proximal and distal inflammation. However, there have been no methods proposed for longitudinal multiple flow FeNO data, and the performance of existing ad hoc two-stage methods has not been evaluated.

Here, we present a novel extension of the U-HB model to longitudinal data (L-U-HB) using a Bayesian implementation of nonlinear mixed effects models. L-U-HB takes as input longitudinal measurements of multiple-flow FeNO on a group of participants to estimate associations of NO parameters with time-varying or time-constant covariates, as well as to quantify within- and between-participant variation. Our work is motivated by longitudinal multiple flow FeNO data collected as part of the Southern California Children's Health Study (CHS)¹⁸. The CHS, originally designed to study impacts of long-term air pollution exposures on children's respiratory health, included repeated measurements of multiple flow FeNO in the most recent cohort.

Methods

First, we introduce the mathematical deterministic model for FeNO in the respiratory tract and then we describe statistical methods for estimating NO parameters from this model using multiple flow FeNO data, including two-stage (TS) approaches and unified (U) approaches for longitudinally assessed multiple flow FeNO.

Deterministic two compartment model for FeNO. Our work is based on the simple steady-state two-compartment model (2CM), which assumes a cylindrically-shaped airway compartment with related NO parameters: C_{aw} , the concentration of NO in the airway tissue (ppb); D_{aw} , the airway tissue diffusion capacity ($\text{pL}\cdot\text{s}^{-1}\cdot\text{ppb}^{-1}$), and an expansile alveolar compartment with related NO parameter C_A , the concentration of NO in the alveolar region (ppb)⁸. Under the 2CM, FeNO (ppb) at the mouth is deterministically related to expiratory flow rate (ml/s) and the three NO parameters quantifying airway and alveolar sources of NO, as shown below:

$$FeNO = C_{aw} + (C_A - C_{aw}) \times e^{-\frac{D_{aw}}{flow}} \quad (1)$$

Estimating NO parameters in the 2CM. The 2CM model for FeNO in Eq. (1) is deterministic and nonlinear. In practice, multiple flow FeNO data is measured with error. Researchers have developed various methods to estimate 2CM NO parameters using multiple flow data from a given participant, typically using linear regression approaches with an underlying linearization assumption, a third order approximation method such as the Högman and Meriläinen algorithm (HMA)^{19,20}, and nonlinear regression which essentially adds an error term to the right hand side of Eq. (1)²¹. Here and in previous work^{10,12}, we use the following fundamental nonlinear statistical model for multiple flow FeNO measured repeatedly across a range of flow rates for a single participant, with maneuvers indexed by k :

$$\log(FeNO_k) = \log\left(C_{aw} + (C_A - C_{aw}) \times e^{-\frac{D_{aw}}{flow_k}}\right) + \varepsilon_k \quad (2)$$

This model formulation includes a “transform-both-sides”²² approach using the natural log to acknowledge the increased variation in error that occurs as flow rate (and hence FeNO concentration) increases while maintaining the interpretability of the 2CM NO parameters. On the logFeNO scale, the error (ε) can be reasonably assumed to be normally distributed. Henceforth, we will refer to a model estimating NO parameters using Eq. 2 with standard nonlinear-least squares software (e.g., “nls” from the nlme package in R) as NLS¹⁰. So far, we have discussed only estimation of NO parameters from a single multiple flow FeNO test session for one participant. When multiple flow FeNO data are assessed longitudinally in a study population the data have three levels of variation: across-participant, within-participant (across visits), and within-visit (across maneuvers).

Estimating associations of covariates with longitudinally assessed NO parameters. *Two Stage (TS) methods.* In the existing literature, most researchers use ad hoc two stage approaches to relate estimated NO parameters to covariates. Two stage methods for cross-sectional studies were discussed in our previous work¹². A typical longitudinal two stage method proceeds as follows. In Stage I, NO parameters for each participant at each visit are estimated via separate models. For example, a separate HMA or NLS model is fit to the multiple flow FeNO data from each participant at each visit. In Stage II, three linear mixed effect models (LMMs) are fit, one for each NO parameter, to relate the longitudinal estimates of the NO parameters to a covariate(s) of interest, denoted generically as X_{ij} . A participant-level random intercept is included in each LMM to account for the within-participant correlation in the longitudinal NO parameter data. Below, we introduce 4 longitudinal two stage (L_TS) methods, differentiated by the name of the method employed in Stage I:

1. L_TS_NLS: Stage I consists of N (participants) \times M (visits) separate NLS¹⁰ models, each NLS model fit to the typically small multiple flow FeNO dataset at that visit, using the natural log transform-both-sides approach discussed earlier.
2. L_TS_HMA: Similarly, Stage I consists of $N \times M$ separate HMA^{19,20} models.
3. L_TS_NLME: Stage I consists of a single longitudinal nonlinear least square mixed effect (NLME) model, an extension of the approach using $N \times M$ separate NLS models, again using the natural log transform-both-sides approach. In the longitudinal NLME, we specified participant-level and visit-level random intercepts for each NO parameter. At each level, these random effects follow a multivariate normal distribution, allowing

- for correlation of NO parameters. For example, participant-level correlation allows for participants with high C_A to also tend to have high C_{aw} . We implemented NLME using the nlme package in R (version 3.1–152)²³.
4. L_TS_HB: Stage I consists of a single longitudinal Hierarchical Bayesian (HB) analog of the longitudinal NLME model, implemented using JAGS (Just another Gibbs sampler)²⁴ similar to the U-HB cross-sectional model in our previous publication¹², but with no covariate X and a partitioning of variance in NO parameters at the participant- and visit-levels through specification of variance–covariance matrices, the population mean of the NO parameters, and the measurement error in Stage I. This model is described in greater detail in the L-U-HB section below, where the model includes X.

All these TS approaches use the same LMM approach in Stage II.

Unified approaches. Unified methods, in contrast to TS methods, simultaneously estimate NO parameters and their associations with the covariate X_{ij} in a single model. In this longitudinal version, we estimated the between/within-participant variation at the same time, which L-TS-NLS and L-TS-HMA were not able to obtain because their estimation in the stage I ignored the grouping effect.

5. L-U-HB

Our novel U-HB model for longitudinal data (U-HBL) has three levels: maneuver, visit, participants, as described below and displayed in Fig. 1:

Level 1: Maneuver

$$\log(\text{FeNO}_{ijk}) \sim N(f(\theta_{ij}, \text{flow}_{ijk}), \sigma_\epsilon^2) \tag{3}$$

$$f(\theta_{ij}, \text{flow}_{ijk}) = \log\left(\exp(\log C_{aw\ ij}) + (C_{A\ ij} - \exp(\log C_{aw\ ij})) \times e^{-\frac{\exp(\log D_{aw\ ij})}{\text{flow}_{ijk}}}\right) \tag{4}$$

In the first level, log FeNO for participant i at visit j and maneuver k is assumed to be normally distributed with a mean that is a function of NO parameters: $\theta_{ij} = (C_{A\ ij}, \log C_{aw\ ij}, \log D_{aw\ ij})'$ and expiratory flow, flow_{ijk} . The variance of the unexplained error in logFeNO, σ_ϵ^2 , was assumed to be the same across flow rates, visits, and participants.

Level 2: Visit (time)

$$\theta_{ij} = A_{\Theta_i} + \beta_{1\Theta} X_{ij} + \alpha_{\Theta_{ij}} \tag{5}$$

In the second level, NO parameters for participant i at visit j (θ_{ij}) are modeled as a linear function of A_{Θ_i} , a vector of participant-level mean NO parameter values for participant i when the covariate $X_{ij} = 0$. Key parameters of interest include $\beta_{\Theta} = (\beta_{C_A}, \beta_{\log C_{aw}}, \beta_{\log D_{aw}})'$, the regression coefficients on X_{ij} . Otherwise unexplained within-participant variation in the NO parameters is represented by the visit-level random intercepts $\alpha_{\Theta_{ij}} = (\alpha_{C_{Aij}}, \alpha_{\log C_{awij}}, \alpha_{\log D_{awij}})$ assumed to have no correlations and follow a multivariate normal distribution (MVN) with variance–covariance matrix $\Sigma_\sigma = \text{diag}(\sigma_{C_{Aij}}^2, \sigma_{\log C_{awij}}^2, \sigma_{\log D_{awij}}^2)$, i.e., $\alpha_{\Theta_{ij}} \sim \text{MVN}(0, \Sigma_\sigma)$.

Level 3: Participant

Priors/Hyperparameters:

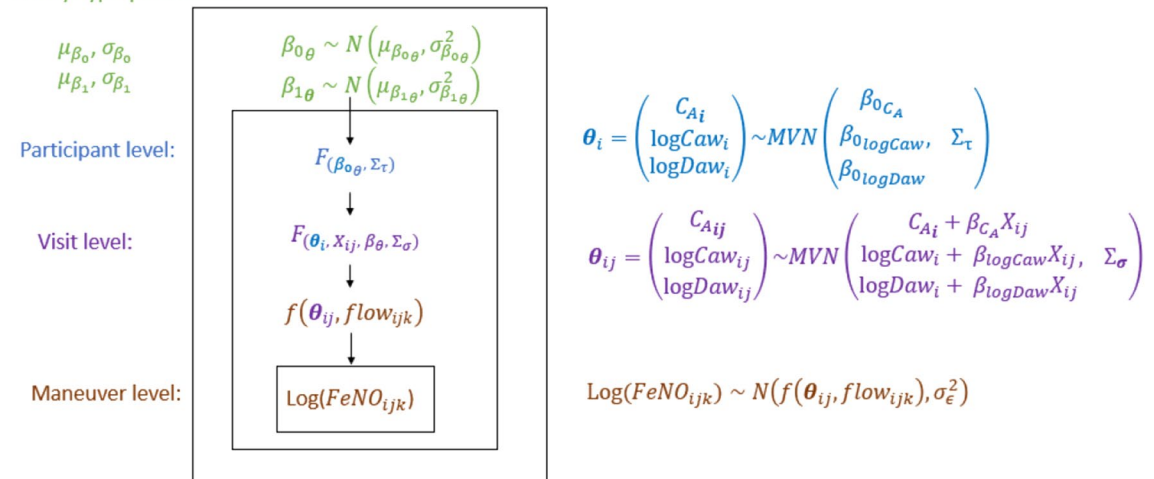


Figure 1. Diagram of the hierarchical model structure relating Longitudinal FeNO measurements at multiple flow rates to NO parameters that are a function of a potential determinant X (e.g., air pollution).

$$A_{\Theta_i} = \beta_{0_{\Theta}} + \alpha_{\Theta_i} \tag{6}$$

In the third level, A_{Θ_i} is decomposed into $\beta_{0_{\Theta}}$, the overall population-mean NO parameters when $X_{ij} = 0$ and the otherwise unexplained between-participant variation in the NO parameters, represented by the participant-level random intercepts $\alpha_{\Theta_i} = (\alpha_{C_A}, \alpha_{\log C_{aw}}, \alpha_{\log D_{aw}})'$, assumed to follow a MVN with variance–covariance matrix Σ_{τ} , i.e., $\alpha_{\Theta_i} \sim MVN(0, \Sigma_{\tau})$.

Levels 2 and 3 are combined in the following equation with two random intercepts: one at the visit level and the other at the participant level.

$$\theta_{ij} = \beta_{0_{\Theta}} + \beta_{1_{\Theta}} X_{ij} + \alpha_{\Theta_i} + \alpha_{\Theta_{ij}} \tag{7}$$

Prior distributions for L-U-HB are specified to be relatively non-informative. We assume the regression coefficients each have independent multivariate normal priors (with I indicate a square identity matrix):

$$\beta_{0_{\Theta}} \sim MVN(\mu_{\beta_0}, I\sigma_{\beta_0}^2) \tag{8}$$

$$\beta_{1_{\Theta}} \sim MVN(\mu_{\beta_1}, I\sigma_{\beta_1}^2) \tag{9}$$

with $\sqrt{\mu_{\beta} = (\mu_{\beta_{C_A}}, \mu_{\beta_{\log C_{aw}}}, \mu_{\beta_{\log D_{aw}}})'}$, a non-informative prior distribution using large variances $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = 10^3$. The random intercept variance–covariance matrix Σ_{ν} is assumed to have a non-informative inverse-Wishart prior distribution with non-informative diagonal matrix $D = \text{diag}(0.001, 0.001, 0.001)$. And the variances of within-participants are sampled from non-informative independent inverse gamma distributions. Finally, the residual variance σ^2 is assumed to have a non-informative inverse-Gamma distribution, *Inv-Gamma* (0.001, 0.001).

The L_U_HB and L_TS_HB's first stage was simulated via JAGS as mentioned above. The simulation process includes an adaptive mode phase (“burn-in”) and a long enough updating phase²⁴ where the adaptive mode is turned off.

6. L-U-NLME

The longitudinal version of the unified NLME model is similar to the cross-sectional one in which the covariate is linked to the mean function for NO parameters, except that it also specifies the variance–covariance matrix for the visit level subgroup. We also specify the diagonal matrix for the visit level variations for our simplified model.

Constraint on C_A . C_A must be non-negative since it represents the concentration, in ppb, of the NO in the alveolar compartment¹². There are similar constraint considerations for C_{aw} and D_{aw} which we satisfy by modeling $\log C_{aw}$ and $\log D_{aw}$ since C_{aw} and D_{aw} to have approximately log-normal population-level distributions. C_A tends to have more of a normal or truncated normal distribution, so a different approach is necessary. In the simulation study data generation step (described below), we discard samples if their C_A was negative. We also enforce the non-negative constraint in the HB models by using a truncated distribution function. We didn't apply the constraint on HMA since our previous papers^{10,12} proved that it had poor performance, probably due to the large number of failed to converge in Stage I. The NLME models implemented in the nlme package in R were fitted without such constraints since there are no readily available constraint options. Constrained versions of TS_NLS proved to be more biased in our previous study¹², thus we implement only the unconstrained NLS here.

Simulation study. We compare the above methods in an extensive simulation study, roughly based on the CHS study design. Each simulated dataset consists of 500 participants with 3 visits each and each visit includes 8 multiple flow FeNO maneuvers (2 each at: 30, 50, 100, and 300 ml/s), which we simulate under a given “scenario” of underlying true associations of the NO parameters with a standard normal covariate X_{ij} (independent across and within participants). For a given scenario, 100 replicate datasets (each of $N \times M = 500 \times 3$) are generated. Data-generating values, shown in Table 1, of population-level mean NO parameters ($\beta_{0_{\Theta}}$), the between-participant variance–covariance matrix ($\Sigma_{\Theta, \nu}$), and the residual variance σ_{ϵ}^2 are based on values estimated in a preliminary

NO parameter	Population-level (across-participant)		Within-participant (across-visit)	Population-level (across-participant)	
	Mean (μ)	Standard Deviation (τ)	Standard Deviation (σ)	Correlation (ρ)	
C_A	1.5	0.45	0.50	$C_A, \log C_{aw}$	0.66
$\log C_{aw}$	3.5	0.65	0.40	$C_A, \log D_{aw}$	−0.38
$\log D_{aw}$	2.5	0.55	0.33	$\log C_{aw}, \log D_{aw}$	−0.35

Table 1. Parameter values used to generate data in the simulation study.

L-TS-NLME analysis of CHS data and are similar to the values in the previous cross-sectional study¹². We set the with-participant covariances to be zero for simplicity.

The different scenarios for the simulation study are described in Table 2. The regression coefficients $\beta_{1\epsilon}$ relating X_{ij} to each NO parameter take a range of values: 0.01, 0.05 to 0.1. In Scenario 1, the reference scenario, all NO parameters have the same association with the covariate X_{ij} ($\beta_{C_A} = \beta_{\log C_{aw}} = \beta_{\log D_{aw}}$), and the value of this association is either 0.01, 0.05 or 0.1.

In this simulation study we compare performance of the methods based on several metrics: percent bias (% bias), 95% Confidence/Credible (CI) length and coverage, power, and type I error. Percent bias is calculated as: (Estimate-True value)/True value for non-zero parameters. The 95% CI length and coverage are calculated using the given 95% confidence interval for frequentist approaches and 95% credible interval (using the returned posterior distribution, the center portion contains 95% of the values) for Bayesian approaches. Primary analyses assume Scenario 1 to be true and use Scenario 1 results to calculate bias, power, 95% CI coverage, 95% CI length while Scenarios 2–4 are used only to calculate type I errors. In secondary analyses, Scenarios 2–4 are used to calculate bias, power, and type I error rate, 95% CI coverage, 95% CI length.

CHS data analysis. We analyzed data from a CHS cohort originally recruited in kindergarten/1st grade²⁵ with FeNO₅₀ assessed at 6 study visits over 8 years (spanning ages 8–16) and multiple flow FeNO assessed at the last 2 study visits, when most children were ages 13–14 and 15–16. The CHS multiple flow FeNO protocol called for 9 maneuvers at each of four expiratory flow rates (3 at 50 ml/s and 2 each at: 30, 100, and 300 ml/s) collected using chemiluminescence analyzers (model CLD88-SP with DeNOx accessory to provide NO-scrubbed air; EcoMedics, Duernten, Switzerland/Ann Arbor, MI, USA) as described in detail elsewhere^{26,27}. FeNO data processing was based on the ATS/ERS guidelines for FeNO at 50 mL·s⁻¹⁷ with a search window based on airway turnover²⁸. Each CHS child participant provided informed assent and a parent/guardian provided informed consent. The CHS data were collected using a protocol approved by the University of Southern California Institutional Review Board, the analyses in this paper were conducted under HS-13-00,150, and all methods were carried out in accordance with relevant guidelines and regulations.

In a previous longitudinal analysis, FeNO₅₀ was found to have a strong positive linear association with height across this age range in children without asthma¹⁷. To complement the previous analysis relating longitudinally assessed FeNO₅₀ to height, here we relate longitudinally assessed NO parameters (from the up to 2 repeated assessments of multiple flow FeNO) to standardized height (population-mean centered: 162.7 cm and population-SD scaled: 8.75 cm). The analyses included 1004 children who never reported a doctor diagnosis of asthma and had multiple flow FeNO data available at both visits. The average number of valid multiple flow maneuvers at the first and second visits were 9.68 and 8.97, respectively.

Results

Simulation study. Computation time was longer for unified methods than for two-stage methods, as expected. The computation time for a given method was similar across scenarios and slightly shorter for larger β coefficients (Supplementary Table 2). Average computation times on a high-performance computing platform (3 CPU, 12 GB memory) for a single simulated dataset (500 participants, 3 visits each, 8 maneuvers per visit) were: 30 h for L_U_HB, 23 h for L_TS_HB, 14.7 min for L_U_NLME, 11.4 min for L_TS_NLME, 4.6 s for L_TS_HMA and 6.2 s for L_TS_NLS. Most methods had reasonable convergence rates (99% for L_TS_HB, 93% for L_U_HB, 94% for L_TS_NLME) except for L_U_NLME (51%). L_TS_NLS and L_TS_HMA converged for all datasets, however Stage I of L_TS_NLS had 29% of participant models fail to converge on average (resulting in the exclusion of these participants' results in Stage II) while L_TS_HMA had only 0.016% failures in Stage I (Supplementary Fig. 2). The following simulation study results are a summary of all available converged results, since the intersection of datasets which converged under all methods is relatively small due to the convergence issue for L_U_NLME.

Figure 2 compares the percent bias and 95% CI interval properties for estimation of β_{C_A} , $\beta_{\log C_{aw}}$, $\beta_{\log D_{aw}}$ across methods. For many methods, percent bias tended to decrease as the effect size of the covariate increased. Among all methods, L_U_HB had the lowest absolute values of percent bias for all three NO parameter associations (all < 4%), however at a small magnitude effect size (true β of 0.01) there was 52% bias for $\beta_{\log D_{aw}}$, equivalent to a 0.052 bias on the original scale. L_U_NLME also had good performance in the subset of datasets where the method converged. Two stage methods (L_TS_HB, L_TS_NLME, L_TS_NLS) tended to have negative percent bias for β_{C_A} , $\beta_{\log D_{aw}}$ and positive percent bias for $\beta_{\log C_{aw}}$. For a given method and effect size, percent bias was smaller for β_{C_A} than for associations with other NO parameters, perhaps because C_A is in the linear part of the 2CM for FeNO. In simulation Scenario 3, when only $\log C_{aw}$ had an association with X , the directions of bias for

	β_{C_A}	$\beta_{\log C_{aw}}$	$\beta_{\log D_{aw}}$
Scenario 1	0.01,0.05,0.1	0.01,0.05,0.1	0.01,0.05,0.1
Scenario 2	0.01,0.05,0.1	0 [†]	0 [†]
Scenario 3	0 [†]	0.01,0.05,0.1	0 [†]
Scenario 4	0 [†]	0 [†]	0.01,0.05,0.1

Table 2. The 4 simulation study scenarios are each repeated at 3 effect sizes and replicated 100 times, for 1200 simulated datasets in total. [†]Cells marked 0 indicate that X had no effect on the corresponding NO parameter.

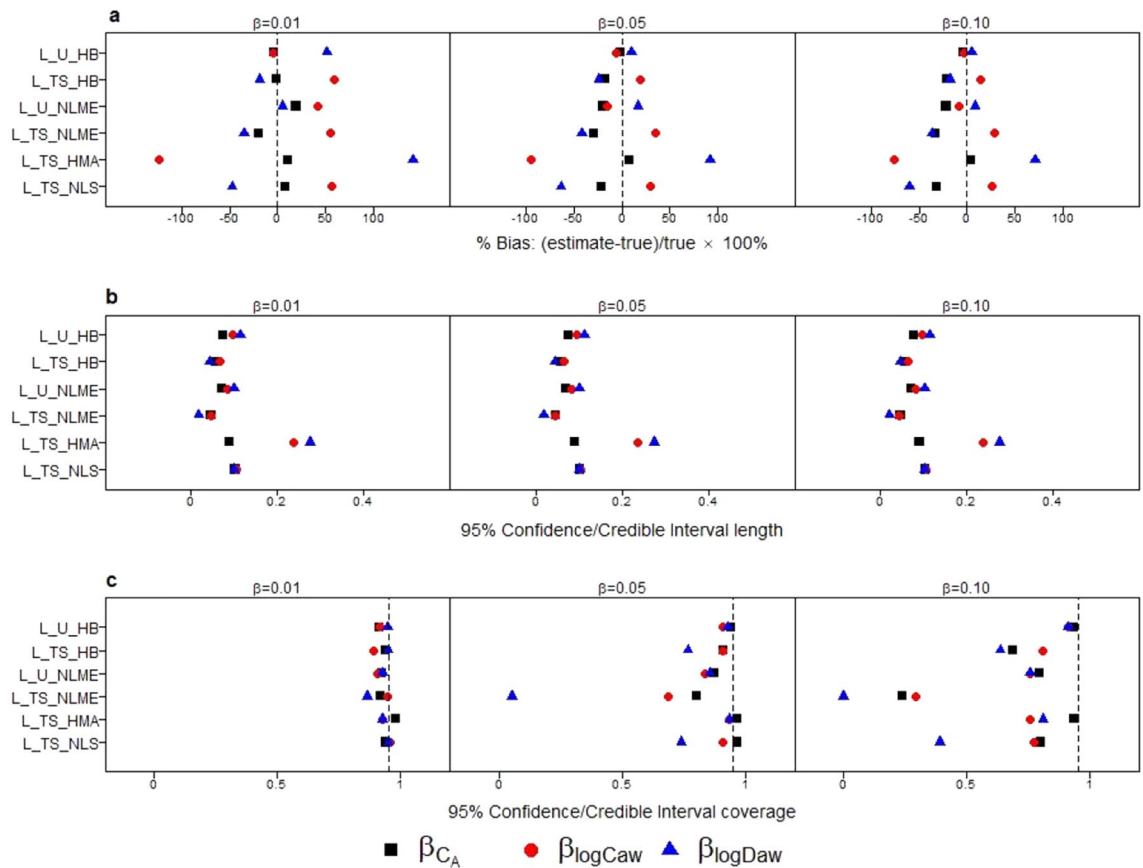


Figure 2. Comparison of method performance in terms of: percent bias (a), 95% CI length (b), and 95% CI coverage (c) for estimating associations with NO parameters (β_{Ca} : black square, $\beta_{\log Caw}$: red circle, $\beta_{\log Daw}$: blue triangle) in simulation study Scenario 1, replicated at 3 different effect sizes ($\beta = 0.01, 0.05$ or 0.1).

two stage methods were different from other scenarios (Supplemental Fig. 1.C1). But L_U_HB and L_U_NLME still had the smallest bias. An alternative version of L_TS_HMA with the constraint that Stage I $C_A > 0$ in Scenario 1 (Supplementary Fig. 3) resulted in $\sim 30\%$ of Stage I estimates being dropped, on average, and patterns in percent bias similar to other two stage methods.

For a given method, the 95% CI lengths (Fig. 2b) were similar across NO parameters and these patterns did not vary based on magnitude of the true β . L_TS_NLME had the shortest 95% CI lengths followed by L_TS_HB. For 95% CI coverage (Fig. 2c), coverage declined as the magnitude of the true β increased, for all methods except L_U_HB. It was the only method to produce 95% CIs with approximately appropriate coverage (91–95%) for all β s. L_U_NLME had slightly larger biases and shorter 95% CI lengths, resulting in 75–80% coverage when the true β was 0.1. The higher coverage for these two unified methods was due to a combination of low bias and longer 95% CI. L_TS_HB also had reasonable coverage, but it differed across NO parameters, ranging from 63 to 81%. For L_TS_NLME, L_TS_HMA, and L_TS_NLS, low coverage was due to a combination of large bias and short CI lengths.

Power curves and Type I error rates for β_{Ca} , $\beta_{\log Caw}$, $\beta_{\log Daw}$ across all simulation scenarios are shown in Supplementary Fig. 1. For simplicity, here Fig. 3 displays a subset of these data: the power for each NO parameter association versus two versions of type I error, based on scenarios 2–4 at the largest magnitude effect size considered (true β of 0.1). Ideally, a method will produce 2 values in the upper left-hand corner of this plot, which indicates high power and low type I error regardless of which another NO parameter had a non-zero association. Indeed, L_U_HB had relatively high power (1.00 for β_{Ca} , 0.97 for $\beta_{\log Caw}$, and 0.89 for $\beta_{\log Daw}$) and low type I error rates from 0.02 to 0.07 for all three NO parameter associations except 0.12 for $\beta_{\log Caw}$ in Scenario 4 where only $\beta_{\log Daw}$ was non-zero. Other methods generally also had good power, but L_TS_NLS had low power for $\beta_{\log Daw}$ and L_TS_HMA had low power for both $\beta_{\log Caw}$ and $\beta_{\log Daw}$. Except for L_U_HB, most methods had inflated type I error for $\beta_{\log Caw}$ when $\beta_{\log Daw}$ was non-zero, or vice versa. For example, L_TS_HB had excellent power for β_{Ca} (0.98) and low type I error (0.02) when $\beta_{\log Caw}$ was nonzero but higher type I error (0.17) when $\beta_{\log Daw}$ was nonzero. This issue became more pronounced for $\beta_{\log Caw}$, where L_TS_HB again had excellent power for $\beta_{\log Caw}$ (1.00) and low type I error (0.01) when β_{Ca} had a nonzero association but very high type I error (0.69) when $\beta_{\log Daw}$ was nonzero. In summary, L_U_HB had high power and low type I error rates for all NO parameter associations while other methods had good power and type I error for β_{Ca} and good power but inflated type I error rates for $\beta_{\log Caw}$ and $\beta_{\log Daw}$. Exceptions to this pattern were L_TS_NLS and L_TS_HMA, both of which had low power for $\beta_{\log Daw}$ due to their large negative biases.

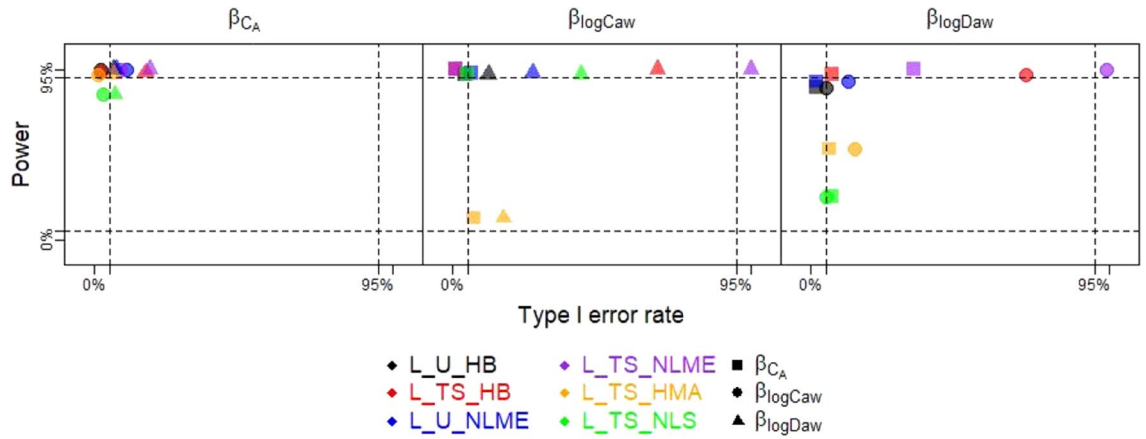


Figure 3. Power and Type I errors for the 6 methods (distinguished by color) for true β of 0.1, with power for a given NO parameter’s association calculated from the simulation scenario where only that association is non-zero (e.g., power for β_{Ca} from Scenario 2 where $\beta_{Ca} = 0.1, \beta_{logCaw} = 0, \beta_{logDaw} = 0$) and Type I error calculated under two scenarios (e.g., Type I error for β_{Ca} under S3: $\beta_{Ca} = 0, \beta_{logCaw} = 0.1, \beta_{logDaw} = 0$ and under S4: $\beta_{Ca} = 0, \beta_{logCaw} = 0, \beta_{logDaw} = 0.1$), with the non-zero NO parameter association denoted by shape.

While the primary focus of the simulation study was on the estimation of $\beta_{Ca}, \beta_{logCaw},$ and $\beta_{logDaw},$ we also studied estimation of random effect variances at the participant and visit levels. Participant-level variances, particularly for C_A and $\log D_{aw},$ tended to be underestimated by most methods, though L-U-HB had the overall lowest bias (Supplemental Fig. 1) Participant-level correlations also tended to be underestimated (Supplemental Fig. 1).

In summary, this simulation study demonstrated that L_U_HB generally had the smallest bias, appropriate 95% CI coverage, and good power with low type I error rates for all three NO parameter associations across scenarios. L_TS_HB, the two-step version of L_HB, had greatly reduced computation time and maintained good power at the expense of introducing some bias, poorer 95% CI coverage at larger magnitude effects, and high type I errors for β_{logCaw} and $\beta_{logDaw}.$ Compared to L_TS_HB, L_U_NLME had similar performance, and even less inflated type I error rates but failed to converge in ~50% of the simulated datasets. Compared to L_TS_HB, L_TS_NLME had bias in the same direction but of larger magnitude, resulting in lower coverage and more inflated type I errors. L_TS_NLS, on average, had ~40% of participants fail to have Stage I estimates (Supplementary Table 1). Despite this, L_TS_NLS performed well for estimation of $\beta_{Ca},$ but for β_{logCaw} and β_{logDaw} had considerable bias, low power and inflated type I error.

CHS data analysis. Applying the 6 methods to a CHS analysis relating NO parameters to height, we observed associations with height that were: positive or null for $C_A,$ positive for $\log C_{aw},$ and negative or null for $\log D_{aw}$ (Fig. 4). The two unified methods (L_U_HB and L_U_NLME) both estimated similar statistically significant associations between height and all three NO parameters. Specifically, from the L_U_HB model we estimated that, a within-participant increase in height of 8.79 cm was associated, on average, with a 0.079 (95% CI: 0.034, 0.125) ppb increase in $C_A,$ a 0.158 (95% CI: 0.106, 0.212) increase in $\log C_{aw},$ and a 0.106 (95% CI 0.044, 0.171) decrease in $\log D_{aw}.$ These latter two estimates are equivalent to a 17% increase in C_{aw} and a 10% decrease in $D_{aw}.$ From the L_U_NLME model, analogous estimates were similar: 0.092 (95% CI: 0.046, 0.138) for $C_A,$ 0.149 (95% CI: 0.104, 0.194) for $\log C_{aw} - 0.104$ (95% CI $-0.157, -0.052$) for $\log D_{aw}.$ Two-stage methods produced lower estimates for β_{Ca} and $\beta_{logCaw},$ and higher estimates for $\beta_{logDaw}.$ Furthermore, two-stage method estimates were all approximately null and non-significant for $\beta_{logDaw}.$ Estimates of β_{Ca} were not statistically significant for L_TS_NLS and L_TS_HMA. There was a clear pattern when comparing a unified method to its

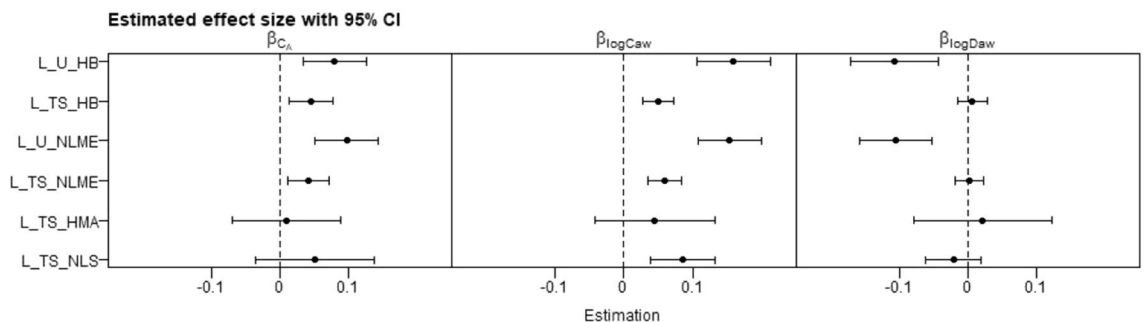


Figure 4. Estimated associations between NO parameters and standardized height in the CHS using 6 methods.

two-stage counterpart (i.e., L_U_HB vs L_TS_HB and L_U_NLME vs L_TS_NLME) in that the unified method had larger magnitude estimates (farther from zero) than their corresponding two-stage versions. The finding that the unified methods produced more significant estimates than two-stage ones (especially the L_TS_NLS and L_TS_HMA), which agreed with results in our simulation. For L_TS_NLS, a Stage I outlier (-1325) for C_A resulted in an extremely wide 95% CI for β_{C_A} in Stage II, so we excluded this outlier for Fig. 4. Stage I convergence failures were observed for 528 and 30 out of 1004 models for L_TS_NLS and L_TS_HMA, respectively.

Discussion

In this paper, we proposed a novel unified hierarchical Bayesian model, L_U_HB, for relating longitudinally assessed NO parameters to covariates, extending our previous cross-sectional unified hierarchical Bayesian model, and performed the first evaluation of the statistical properties of various two-stage methods for longitudinal analysis of NO parameters. In a simulation study, L_U_HB performed well for estimating associations of NO parameters with covariates, with small bias, appropriate 95% CI coverage, good power, and low type I error rates. The two-step version, L_TS_HB, had greatly reduced computation time and maintained good power at the expense of introducing some bias, poorer 95% CI coverage at larger magnitude effects, and high type I errors for $\beta_{\log C_{aw}}$ and $\beta_{\log D_{aw}}$. The other unified method L_U_NLME had good performance when it converged, but it had serious convergence issues. Other two-stage methods had drawbacks in terms of bias, inflated type I error rates, etc.

In a previous simulation study comparing the performance of methods estimating NO parameter associations with a covariate in a cross-sectional study¹², U_HB had the best performance across all simulation scenarios, similar to our findings in this longitudinal study. L_U_NLME had the second-best performance across all three NO parameters in the longitudinal study, while its cross-sectional version (U_NLME) had large bias in estimating β_{C_A} . L_TS_NLS also had much better performance in estimating β_{C_A} and $\beta_{\log C_{aw}}$ in the longitudinal study than TS_NLS for the cross-sectional study but still had large bias for $\beta_{\log D_{aw}}$. Both L_TS_NLME and the cross-sectional TS_NLME had large bias and poor coverage.

Limitations to L_U_HB include computation time. L_U_HB had superior statistical properties to many competitor methods, albeit at additional computational expense. The cross-sectional U_HB model had an average computation time of 5.5 h for $N = 1000$ participants (ref) while for the longitudinal version (L_U_HB) average computation time was 30 h for $N = 500$ participants, 3 visits each. While the longer L_U_HB computation time was burdensome in a simulation study with 1000 s of datasets, it is less of an issue when analyzing a single dataset. However, note that the computational cost will increase as more covariates are added, or a more complex model is used in Stage II. The other competitive method was L_U_NLME, which had the second-best estimation performance and ran faster, but it had poor convergence. Our results suggest that for further applications in which there are more variables or more complex models, where the unified model may become computational intractable, a two-stage version of hierarchical Bayesian model will have reasonably good performance and therefore be used in iterative model building, with a single run of L_U_HB for the final results, using L_TS_HB estimates as starting values to speed convergence. Another issue which should be raised is that any unified estimation framework which simultaneously estimates NO parameters and their associations with covariates will produce NO parameter estimates dependent on the covariates included. Here, our primary interest was in the estimated associations rather than the NO parameters themselves, so this drawback was outweighed by the improved performance in estimating associations.

For biological plausibility, we constrained C_A to be non-negative. In our previous cross-sectional study, we encountered a problem constraining C_A to be non-negative in the standard JAGs software. We solved it by sampling $\log C_{aw}$ and $\log D_{aw}$ as they were bivariate normal distributed, and then sample zero-truncated C_A conditioned on them. In that case, we sampled the variances and correlations one by one and set boundaries for the last correlation to ensure a 3×3 positive definite variance covariance matrix. The same solution was also used for L_U_HB and L_TS_HB when we set NO parameters to be correlated in both levels. But in our simplified model which assumed no correlations in the visit level, the randomness and constraint on C_A were easily specified separately without considering the validity of the variance-covariance matrix.

When applying L_U_HB to study the association between height and NO parameters using longitudinal multiple flow FeNO data on healthy schoolchildren in the CHS, we found positive associations of C_A and $\log C_{aw}$ with height and a negative association of $\log D_{aw}$ with height. Had we applied only L_TS_HMA or L_TS_NLS methods, we would have failed to detect associations of height with C_A or $\log D_{aw}$. Our findings add to the limited literature on associations of NO parameters with height/age. A previous analysis using longitudinal FeNO₅₀ data from the same cohort over a longer follow-up period, from ages 8–16, found that FeNO₅₀ increased approximately linearly with height and FeNO₅₀ increased nonlinearly with age¹⁷. Limited cross-sectional data on trends in NO parameters by age (for participants less than 20 years old) suggests non-significant increases of D_{aw} and C_{aw} but a decrease in C_A though some influential values for the oldest participants in the sample may have impacted these results²⁹. Additionally, differences with our findings may be due to a different age range or due to the difference between a cross-sectional (between-person) versus longitudinal (within-person) design.

There are several directions for future work. To reduce the computation time observed when implementing L_U_HB in JAGS using Gibbs sampling, we could explore alternative Bayesian MCMC software such as RStan which uses Hamiltonian Monte-Carlo. Several components of the model implementation appeared to affect convergence rates, such as the number of parameters or the length of adaptation phase. The length of adaptation was the most important factor, but adaptation is in itself a computationally intensive operation. Given the computational costs involved, we chose to base our simulation study on smaller dataset sizes (500 participants, 3 visits each) and simplify our model so that it could converge within 2 days on average. The simplified version of our longitudinal model ignored the correlation between the NO parameters within the participant level (V0

model). L_U_HB and L_TS/U_NLME were able to specify the diagonal matrix for visit-level variation. While the NLS and the HMA approaches fitted the FeNO models for each observation, thus ignoring any correlations between or within participants. In the CHS data analysis, we only selected participants who have visited both in year 8 and 10 as a more direct comparison to the simulation study, but unified models are capable of handling unbalanced data.

In conclusion, in this paper we presented a longitudinal extension of the unified hierarchical Bayesian model for analyzing nonlinear data (e.g., FeNO data). Despite the long computation time required for achieving convergence, L_U_HB had the best performance estimating covariate coefficients as well as variance–covariance components. The two-stage analog, L_TS_HB, served as a reasonable alternative to explore initial versions of more complicated models.

Data availability

Due to limitations in the original consent forms and HIPAA requirements, data from the CHS cannot be freely available in the public domain. However, we are committed to sharing the data and results acquired as part of this study. The CHS has a process in place for data sharing that involves approval of proposals by a Data Sharing Committee. Investigators who want access to data will be required to submit a research protocol, which will be reviewed by the CHS Health Data Release Committee and the USC IRB. Recipients must agree to security policies. Please send requests to access this dataset to Dr. Sandrah Eckel (eckel@usc.edu).

Received: 29 March 2022; Accepted: 16 March 2023

Published online: 01 April 2023

References

- Jain, N. & Hill, J. L. The National Heart Lung and Blood Institute guidelines finally say “yes” to fractional exhaled nitric oxide. *Ann. Allergy Asthma Immunol.* **128**(4), 348–349 (2022).
- Khatri, S. B. *et al.* Use of fractional exhaled nitric oxide to guide the treatment of asthma: an official American thoracic society clinical practice guideline. *Am. J. Respir. Crit. Care Med.* **204**(10), e97–e109 (2021).
- Dweik, R. A. *et al.* An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. *Am. J. Respir. Crit. Care Med.* **184**(5), 602–615 (2011).
- La Grutta, S., Ferrante, G., Malizia, V., Cibella, F. & Viegi, G. Environmental effects on fractional exhaled nitric oxide in allergic children. *J. Allergy* **2012**, 1–6. <https://doi.org/10.1155/2012/916926> (2012).
- Scarpa, M. C., Kulkarni, N. & Maestrelli, P. The role of non-invasive biomarkers in detecting acute respiratory effects of traffic-related air pollution. *Clin. Exp. Allergy* **44**(9), 1100–1118 (2014).
- Oliver, A. C. *et al.* Effects of reduced nicotine content cigarettes on fractional exhaled nitric oxide and self-reported respiratory health outcomes among smokers with psychiatric conditions or socioeconomic disadvantage. *Nicotine Tob. Res.* **24**(1), 135–140 (2022).
- American Thoracic Society. European Respiratory Society. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *Am. J. Respir. Crit. Care Med.* **171**, 912–930 (2005).
- George, S. C., Hogman, M., Permutt, S. & Silkoff, P. E. Modeling pulmonary nitric oxide exchange. *J. Appl. Physiol.* **96**(3), 831–839. <https://doi.org/10.1152/jappphysiol.00950.2003> (2004).
- Roy, K. *et al.* Use of different exhaled nitric oxide multiple flow rate models in COPD. *Eur. Respir. J.* **29**(4), 651–659 (2007).
- Eckel, S. P. *et al.* Estimation of parameters in the two-compartment model for exhaled nitric oxide. *PLoS One* **9**(1), e85471 (2014).
- Karvonen, T. *et al.* Comparison of feasibility and estimates of central and peripheral nitric oxide parameters by different mathematical models. *J. Breath Res.* **11**(4), 047102 (2017).
- Weng, J. *et al.* Hierarchical Bayesian estimation of covariate effects on airway and alveolar nitric oxide. *Sci. Rep.* **11**(1), 17180 (2021).
- Karrasch, S. *et al.* Accuracy of FENO for diagnosing asthma: a systematic review. *Thorax* **72**(2), 109–116 (2017).
- Petsky, H. L., Kew, K. M., Turner, C. & Chang, A. B. Exhaled nitric oxide levels to guide treatment for adults with asthma. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD011440.pub2> (2016).
- Silkoff, P. E. *et al.* Longitudinal stability of asthma characteristics and biomarkers from the Airways Disease Endotyping for Personalized Therapeutics (ADEPT) study. *Respir. Res.* **17**, 43 (2016).
- Kharitonov, S. A. *et al.* Reproducibility of exhaled nitric oxide measurements in healthy and asthmatic adults and children. *Eur. Respir. J.* **21**(3), 433–438 (2003).
- Garcia, E. *et al.* Patterns and determinants of exhaled nitric oxide trajectories in schoolchildren over a 7-year period. *Eur. Respir. J.* **56**(1), 2000011 (2020).
- Linn, W. S. *et al.* Exhaled nitric oxide in a population-based study of Southern California Schoolchildren. *Respir. Res.* **10**, 28 (2009).
- Hogman, M. & Merilainen, P. Extended NO analysis in asthma. *J. Breath Res.* **1**(2), 024001 (2007).
- Hogman, M. *et al.* Extended NO analysis applied to patients with COPD, allergic asthma and allergic rhinitis. *Respir. Med.* **96**(1), 24–30 (2002).
- Silkoff, P. E. *et al.* Airway nitric oxide diffusion in asthma: role in pulmonary function and bronchial responsiveness. *Am. J. Respir. Crit. Care Med.* **161**(4 Pt 1), 1218–1228 (2000).
- Latif, A. H. & Gilmour, S. G. Transform-both-sides nonlinear models for in vitro pharmacokinetic experiments. *Stat. Methods Med. Res.* **24**(3), 306–324 (2015).
- Pinheiro, J., *et al.* in: *Nlme: Linear and Nonlinear Mixed Effects Models*. R package version 31-110, vol. 3 pp. 1–113 (2013).
- Plummer, M. in: *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. (2003).
- McConnell, R. *et al.* Traffic, susceptibility, and childhood asthma. *Environ. Health Perspect.* **114**(5), 766–772 (2006).
- Linn, W. S. *et al.* Multiple-flow exhaled nitric oxide, allergy, and asthma in a population of older children. *Pediatr. Pulmonol.* **48**(9), 885–896 (2013).
- Linn, W. S. *et al.* Extended exhaled nitric oxide analysis in field surveys of schoolchildren: a pilot test. *Pediatr. Pulmonol.* **44**(10), 1033–1042 (2009).
- Puckett, J. L. *et al.* Impact of analysis interval on the multiple exhalation flow technique to partition exhaled nitric oxide. *Pediatr. Pulmonol.* **45**(2), 182–191 (2010).
- Högman, M. *et al.* Effects of growth and aging on the reference values of pulmonary nitric oxide dynamics in healthy subjects. *J. Breath Res.* **11**(4), 047103 (2017).

Author contributions

J.W. developed the code and conducted the simulation study and data analysis with the guidance of S.P.E. and P.M. N.M. contributed to early phases of code development. J.W. and S.P.E. wrote the manuscript with input from P.M., N.M., W.J.G., and F.D.G. S.P.E. conceived the study and was in charge of overall direction and planning, with input from P.M., W.J.G., and F.D.G.

Funding

This authors funded by National Institute of Environmental Health, Sciences (R01ES027860), Southern California Environmental Health Sciences Center (P30ES007048).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-31774-7>.

Correspondence and requests for materials should be addressed to S.P.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023