Published in final edited form as:

Commun Stat Theory Methods. 2018 ; 47(22): 5418–5434. doi:10.1080/03610926.2017.1395045.

Joint modeling of multivariate hearing thresholds measured longitudinally at multiple frequencies

Mulugeta Gebregziabher^a, Mark A. Eckert^b, Lois J. Matthews^b, Abeba A. Teklehaimanot^a, and Judy R. Dubno^b

^aDepartment of Public Health Sciences, Medical University of South Carolina, Charleston, USA

^bDepartment of Otolaryngology, Medical University of South Carolina, Charleston, USA

Abstract

Pure-tone thresholds are used to estimate hearing acuity and, when measured longitudinally, can characterize age-related changes in hearing. Measured at multiple-frequencies, multiple-irregular time points, for right and left ears, these longitudinal studies of age-related hearing loss produce data of inherent complexity due to: 1) multivariate outcomes at different frequencies; 2) longitudinal measurements taken at subject-specific time intervals; and 3) inter-ear correlations due to clustering and nesting. To address limitations in existing methods, we propose a multivariate generalized linear mixed model (mGLMM) and assess its performance. We demonstrate its application using a unique dataset from a cohort study of age-related hearing loss.

Keywords

correlated data; longitudinal hearing thresholds; mGLMM; sGLMM multivariate outcomes

1. Introduction

To assess age-related changes in hearing, thresholds for pure-tone signals in right and left ears are measured at lower and higher frequencies over multiple years. This results in a complex, multivariate longitudinal dataset useful for characterization of changes in hearing over time as a function of age, gender, their interactions and other covariates. The main complexities include nesting by ear, within subject correlations of responses due to longitudinal measurements and correlations between the multiple outcomes and data imbalance due to missingness.

A common method to analyze hearing thresholds is to fit each measure at each frequency and ear separately using a linear mixed model (Echt et al., 2010; Kiely et al., 2012; Lee et al., 2005; Morrell and Brant, 1991; Park et al., 2010). Some studies have used general estimating equations (Wiley et al 2008) and ANCOVA (Dullemeijer et al., 2010). However, these univariate methods are limited because they ignore the intercorrelation of threshold

measures at multiple frequencies and between left and right ears. They also do not answer the key research questions related to changes in hearing (Verbeke et al., 2014).

Advanced modeling approaches have been developed and assessed to address the limitations of standard univariate approaches (Davidov and Rosen, 2011; Fieuws and Verbeke, 2004, 2006; Rosen and Davidov, 2012; Verbeke et al., 2014; Verbeke, Spiessens and Lesaffre, 2001), but some lead to loss of efficiency (Fieuws and Verbeke, 2004, 2006) and none deal with inter-ear correlation, nesting, and right-left ear differences. Ignoring any right-left ear differences among individuals can lead to incorrect conclusions (Faraway, 2004). For these reasons, methods that use average thresholds (Dullemeijer et al., 2010) or select thresholds from the better or worse ear to represent the results (Fieuws and Verbeke, 2004; Kiely et al., 2012; Park et al., 2010) are problematic. Davidov and Rosen (Davidov and Rosen, 2011; Rosen and Davidov, 2012) focused on the assumption of order restriction in hearing loss over increasing frequency and age, but did not account for nesting by ear. Treating thresholds from the two ears as separate outcomes (Davidov and Rosen, 2011; Rosen and Davidov, 2012; Wiley et al., 2008) increases the dimension of the response (i.e., thresholds at 11 frequencies result in 22 outcomes) leading to model complexity and computational challenges.

A review of analysis of multivariate longitudinal data has been given recently in (Verbeke et al., 2014). They succinctly provide description of the four main model families (marginal models, conditional models, models for association between latent evolutions, models for evolutions of latent variables) for multivariate longitudinal data with their pros and cons. They discuss the application of these approaches and illustrate them using a few data examples. However, so far there is gap in multivariate GLMM methods that do not force order restriction, deal with nesting and allow for global and accurate inference about longitudinal changes in hearing, a key research question in hearing studies. In this paper, we discussed additional complexities of these types of data that are not covered in the review and outline methods of dealing with these complexities via mGLMM. Thus, the goal of this manuscript was to develop and test a unified mGLMM approach using hearing thresholds at mid-to-higher frequencies that deal with the following constraints and limitations of existing work: (1) bias due to ignoring inter-ear correlation from nesting and clustering of responses by ear (2) loss of precision from using sGLMM; (3) existing inconsistencies in data analyses from two ears; and (4) lack of global test for covariates effect on multiple outcomes in this data setting.

The mGLMM approach may be preferred to avoid the inefficient use of data by sGLMM and hence to minimize bias and increase power (Davidov and Rosen, 2011; Fieuws and Verbeke, 2004; Rosen and Davidov, 2012). This approach also allows for a global test of covariate effects on multiple outcomes, which is a key research question of hearing studies. We demonstrate in the current analysis that the mGLMM approach can fill the existing gap in the analysis of hearing loss data.

The methods examined in this study further build on prior methods for jointly modeling multiple cost outcomes (Gebregziabher et al., 2013; Liu et al., 2008), with a focus on understanding the underlying mechanisms of multiple outcomes as a function of covariates.

Using a joint modeling approach in the GLMM framework, we examined age-related changes in hearing while accounting for the interdependence among threshold outcomes at 11 mid-to-higher frequencies and correlation due to nesting and clustering of observations by ear. Variations in mGLMM modeling were considered to examine the impact of varied model complexity and correlations across multiple outcomes and within the longitudinal measures.

Some promising contributions of this analysis include: (1) a unified theoretical framework for the analysis of multivariate hearing thresholds from both ears measured longitudinally assuming homogenous and heterogeneous residual variance structures, (2) comparative assessment of six variations of mGLMM, (3) the application and use of mGLMM for global tests of covariate effect on multiple outcomes, and (4) SAS code that can be used to implement mGLMM.

2. Data Example

In the longitudinal study of age-related hearing loss at the Medical University of South Carolina (MUSC), subjects 18 years of age and older and in good general health are recruited through advertisements and subject referral. At the time of entry into the study, subjects' ages ranged from 18 to >89 years of age, with a mean age of 61.4 years. Subjects were excluded if there was evidence of conductive hearing loss or active otologic or neurologic disease. From the total of 1475 individuals who had been screened for participation at the time of the current analysis, 1357 persons (747 females, 610 males) were enrolled (92%).

To retain a consistent number of subjects actively involved in the longitudinal study, approximately 50 new subjects have been enrolled each year since the start of the study in 1987. Subjects have voluntarily withdrawn from the study due to non-study related illnesses or death, moving from the area, no longer perceiving a benefit of participation, and increased time constraints. Subjects have been discontinued from the study due to difficulty scheduling, non-age-related changes in hearing or otologic or neurologic conditions, and poor test reliability. For each subject, the number of years from entry to cohort to test date (for each visit) was computed and the corresponding threshold was recorded. Of the 1,357 who have participated in the program, longitudinal data covering at least a 3-year period were available from 884 subjects.

Subjects are scheduled approximately once per month for a total of three to six visits to complete protocol that includes pure-tone air conduction thresholds at conventional audiometric frequencies (0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz) and extended-higher frequencies (9, 10, 11, 12, and 14 kHz). Other measures of auditory function include speech recognition in quiet and noise, assessment of middle ear function, otoacoustic emissions, upward and downward spread of masking, and auditory brainstem responses. The protocol also include a battery of cognitive tests, clinical blood chemistries and self-assessment of hearing handicap among other things. After completion of this protocol, subjects are scheduled annually to update their contact information, medical and hearing histories, prescription and nonprescription medication information, and for measurement of pure-tone thresholds at

conventional frequencies and speech recognition in quiet. To obtain longitudinal data, a portion of the protocol is repeated every 2 to 3 years with data obtained at time intervals: baseline, [1,3), [3,6), [6,9), [9–12), [12,15) years, and so on. For measurement of pure-tone thresholds (in dB HL), estimates are obtained using a clinical audiometer with a 5-dB step size. If no response is obtained at the audiometer's maximum output, the following values are assigned based on clinical guidelines: 115 dB HL (for 1 to 6 kHz), 105 dB HL (for 8 kHz), and 115 dB SPL (for 9 to 14 kHz).

3. Statistical Model and Estimation of Parameters

We examined a multivariate linear mixed model approach to characterize age- and gender-related changes over time in hearing loss by jointly studying (1) the effect of covariates such as age and gender on all outcomes (hearing loss across frequencies) simultaneously, (2) how evolutions and trajectories of multiple longitudinal thresholds evolve over time, and (3) the strength of association between the evolutions of all outcomes. Each hearing loss outcome was modeled via a mixed effects model with random slope and intercept; the correlation between hearing loss outcomes at different frequencies was incorporated by specifying a joint distribution of their corresponding random effects under both homogenous and heterogeneous residual variance structure. The random effects could be shared by the different outcomes leading to what are commonly called shared random effect joint models or they could be assumed to take different values and have a joint multivariate distribution. These models are capable of accommodating the correlation within each subject's measurements over time, the correlation among the multiple outcomes from each subject, and allow for better description of the individual differences (Gibbons et al 2010). Also detailed below are how the models account for the nesting and inter-ear correlation due to clustering. Thus, the proposed mGLMM approach uses joint modeling to address analytical constraints on characterizing age- and gender-related threshold changes at each frequency as well as globally for all frequencies while accounting for the correlation among the repeated measures and among the multiple outcomes and nesting (clustering) by ear. The global effect of covariates on multiple outcomes (in our case whether the effect, for example, gender for all frequencies) is tested via a type-III F-test. For each frequency we report the estimates of the trajectory of associations of hearing loss with age, gender, age by gender interaction and the corresponding variations in these estimates.

3.1. Statistical model

Generalized Linear Mixed Model (GLMM): Let $y_i^k = (y_{ijl}^k, \dots, y_{ijm}^k)$ denote a vector of responses for each outcome where y_{ijl}^k is the hearing threshold level in decibels for subject i ($i=1, \dots, n$) at ear j at the l th time ($l=1, \dots, m$) measured at each of the K ($k=1, \dots, K$) hearing frequencies. Suppose $y_i = (y_i^1, \dots, y_i^K)$ is a matrix of the responses and $e_i = (e_i^1, \dots, e_i^K)$ is corresponding matrix of random error from each of the K outcomes. The generalized linear mixed model (GLMM) for each Y_i^k can be given as,

$$E(Y_i^k | X_i, Z_i) = g^{-1}(X_i \beta^k + Z_i b_i^k), \quad (1.1)$$

where $b_i^k \sim N(0, G^k)$ and $X_i = (1, \text{time}, \text{age}, \text{gender}, \text{gender} * \text{age})^T$ and Z_i represent vectors of fixed and random effect covariates, respectively, β^k is the fixed effects parameter, and the G-side covariance matrix G^k is a $q \times q$ covariance matrix for b_i^k , where q is the dimension of the random effects vector. If b_i^k is a vector of random intercept, slope and ear, then $Z_i = (1, t_i, e_i)^T$ and G^k is a 3×3 covariance matrix (Diggle et al., 2002; Fitzmaurice, Laird and Ware, 2004). The dimension of Z_i could increase if a quadratic term in time or another random effect variable is added to the model. Combined with the R-side covariance matrix (a diagonal variance covariance matrix of the residuals with dimension $m \times m$), the G-side matrix leads to a non-diagonal covariance matrix ($V^k = ZG^kZ' + \sigma_k^2 I_m$) that captures the within and between subjects variation within each threshold and nesting by ear.

Multivariate GLMM (mGLMM): When the responses are multivariate longitudinal outcomes (as in the case of the hearing threshold data collected at K frequencies), the GLMMs for each Y_{ijl}^k , are linked through a shared random effect or a joint distribution of the random intercepts and slopes to form mGLMM. These allow for modeling unobserved heterogeneity via variance components for subject and cluster level effects, the correlation of the repeated measures and the multiple outcomes. Without loss of generality, the unified framework is outlined below for the special case of Y_{ijl}^k coming from Gaussian distribution ($Y_{ijl}^k | X_i, Z_i = X_i \beta^k + Z_i b_i^k + e_{ijl}^k$, where $b_i^k \sim N(0, G^k)$ and $e_{ijl}^k \sim N(0, \sigma_k^2)$). Extensions to other distributions such as mixed effects logit or Poisson is straightforward.

The variations of mGLMM are described as special cases of the heterogeneous separate random intercept (hSPRI) and random slope (hSPRIS) model. Assuming the random effects to have identical impact on all outcomes (strong assumption), the model is given as,

$$\begin{aligned}
Y_{ijl}^1 | b_i^1 &= \beta_0^1 + b_{0i}^1 + (b_{1i}^1 + \beta_1^1)t_i + \beta_2^1 x_i + b_{2l}^1 e_{il}^1 + \varepsilon_{ijl}^1 \\
Y_{ijl}^2 | b_i^2 &= \beta_0^2 + b_{0i}^2 + (b_{1i}^2 + \beta_1^2)t_i + \beta_2^2 x_i + b_{2l}^2 e_{il}^2 + \varepsilon_{ijl}^2 \\
&\vdots \\
Y_{ijl}^K | b_i^K &= \beta_0^K + b_{0i}^K + (b_{1i}^K + \beta_1^K)t_i + \beta_2^K x_i + b_{2l}^K e_{il}^K + \varepsilon_{ijl}^K
\end{aligned}$$

$$b_i = \begin{pmatrix} b_i^1 \\ b_i^2 \\ \vdots \\ b_i^K \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} b_{0i}^1 \\ b_{1i}^1 \\ b_{2l}^1 \end{pmatrix} \\ \begin{pmatrix} b_{0i}^2 \\ b_{1i}^2 \\ b_{2l}^2 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} b_{0i}^K \\ b_{1i}^K \\ b_{2l}^K \end{pmatrix} \end{pmatrix} \square MVN(0, G), \varepsilon_i^k = \begin{pmatrix} \varepsilon_{il1}^k \\ \varepsilon_{il2}^k \\ \vdots \\ \varepsilon_{ilm}^k \end{pmatrix} \sim MVN(0, R^k) \quad (1.2)$$

where b_i^k is a vector of random effect terms that each have covariance matrix G^k . In the *mGLMM*, the vector b_i is assumed to have a multivariate normal with mean zero and covariance matrix G (3K by 3K dimension). The R-side variance for the k th outcome can be assumed to be homogenous ($\sigma_k^2 I_m$) or heterogeneous ($\sigma_k^2 I_m$) or it can further be grouped by ear to account for ear level heterogeneity. The heterogeneous variance can be modeled in different ways. For example, for each outcome, it could be a diagonal matrix with different variance components for the m repeated measures leading to sGLMM. On the other hand, the outcomes could be grouped by the similarity of their variation to reduce the dimension of the G-side and R-side covariance matrix. The hSPRI and hSPRIS reported in the results section, for instance, are based on four groupings of the 11 frequency thresholds (1–3 kHz; 4–6 kHz; 8–10 kHz; 11–14 kHz) based on similarities of the individual profiles of HL. The overall R-side variance is a block-diagonal matrix with entries R^k . Combining the R-side and G-side covariance matrices, we get a non-diagonal covariance matrix ($V = ZGZ' + R$) that captures the within subject, between subjects, nesting by ear and between outcomes variation.

Six variations of mGLMM that range in complexity from the separate analysis of outcomes to the very restrictive shared random intercept model are described below.

1. The most complex model assumes separate random effects for each outcome and a heterogeneous diagonal R-side matrix leading to results equivalent to separate analysis of each outcome using GLMM. The only difference is that this provides

a global F-test for the effect of each of the covariates on the multiple outcomes. We can have this as

- a. **HSPRI**- separate random intercepts with **heterogeneous** R-side variance (which is block diagonal matrix with elements $R^k = \sigma^2_{[k]} I_m$)
 - b. **HSPRIS**- separate random intercept and random slope with **heterogeneous** R-side variance (which is block diagonal matrix with elements $R^k = \sigma^2_{[k]} I_m$)
2. The assumption of heterogeneous R-side matrix can be relaxed by grouping outcomes with similar variability profile as it is customarily done in GLMM (Fitzmaurice et al., 2004). For example, in our hearing loss study, we considered four groupings of the 11 frequency thresholds 1kHz-3kHz, 4kHz-6kHz, 8kHz-10kHz, and 11kHz-14kHz based on similarities of the individual profiles of hearing loss. We can have this as
 - a. **HSPRI**- separate random intercepts with **heterogeneous** R-side variance (which is block diagonal matrix with elements $R^k = \sigma^2_{[k^*]} I_m, k^* < k$)
 - b. **HSPRIS**- separate random intercept and random slope with **heterogeneous** R-side (which is block diagonal matrix with elements $R^k = \sigma^2_{[k^*]} I_m, k^* < k$)
3. The assumption of heterogeneous R-side matrix can be restricted to homogenous R-side as is done in GLMM (Fitzmaurice et al., 2004). We can have this as
 - a. **SPRI**- separate random intercepts with **homogenous** R-side variance (which is block diagonal matrix with elements $R^k = \sigma^2 I_m$)
 - b. **SPRIS**- separate random intercept and random slope with **homogenous** R-side variance (which is block diagonal matrix with elements $R^k = \sigma^2 I_m$)
4. A further restriction is to assume perfect correlation among the random intercepts/slopes resulting in shared random effects model (Verbeke et al., 2014). We can have this as
 - a. **SHRI**-shared random intercept with **homogenous** R-side variance
 - b. **SHRIS**-shared random intercept and slope model with **homogenous** R-side

The key underlying assumptions are: (1) Conditional independence where the responses are independent given the random effects **b_i**. (Fieuws, Verbeke and Molenberghs, 2007; Fitzmaurice et al., 2004). (2) The covariance matrix V reflects the dependence structure of responses within each frequency (due to repeated measurements) and across multiple frequencies. (3) The elements of V can be used to account for (i) inter-outcome correlation

at time j (which can be made to vary by occasion), (Park et al.) intra-outcome correlation for each subject (which can be made to vary by outcome) and (Park et al.) cross-correlation for outcome type k at time j and outcome type k' at time j' (which can be made to vary by occasion and outcome). The joint model links the GLMM models for each outcome by assuming a joint statistical distribution for the random coefficients (intercept and slopes).

3.2. Modeling left and right ears

As described earlier, the common approaches to modeling data from two ears of an individual include averaging the thresholds from the two ears, selecting a better or worse ear (or right or left ear) to represent the results for the individual, or treating the responses from the two ears as outcomes leading to doubling of the dimension of the response variable. None of these accounts for the inter-ear correlation due to nesting by ear on the overall error variance of the hearing thresholds, which depends on whether there are right-left ear differences between individuals. Here, we propose a unified approach that allows for data from both ears to be modeled, accounting for inter-ear correlation due to nesting and clustering of observations by ear. To implement this we include a shared random effect, b_{2j} or separate random effects for each frequency b_{2j}^k as nested random effect. This leads to a G matrix of dimension 33×33 , 11×11 , 3×3 and 2×2 in the SPRIS, SPRI, SHRIS and SHRI respectively. Although pooling of ear-level data has been the norm, the pooling decision can be made empirically by testing whether the variance of b_{2j}^k is different from zero via LRT.

3.3. Parameter Estimation and Inference

The parameters are identified as the values that optimize the objective function given below. For Gaussian outcomes, this function can be a maximum likelihood (ML) or restricted maximum likelihood (REML) function. This can be implemented in the SAS Proc GLIMMIX (SAS Institute Inc.). This is a generalization of what can be done using Proc MIXED (Gao Feng et al., 2003; Thiebaut Rodolphe et al., 2002). The MLE is obtained by solving the likelihood under the assumption of conditional independence of $(y_i^I, \dots, y_i^K)$ given b_i using the Laplace or QUAD approximation. Assuming $\text{var}(b_i) = G$ and $\text{var}(\epsilon_i) = \Sigma_i$, the log-likelihood is given as,

$$l = -\frac{1}{2} \sum_{i=1}^n \left\{ m_i \log(2\pi) + \log |V_i| + (y_i - x_i \beta)^T V_i^{-1} (y_i - x_i \beta) \right\} \quad (1.3)$$

where $V_i = Z_i G Z_i' + \Sigma_i$ and both G and $\Sigma_i = \text{diag}\{R^I, \dots, R^K\}$ are assumed to be unknown.

Empirical Bayes estimation is used to predict the b_i and marginal maximum likelihood to estimate the elements of the variance-covariance matrix V . For the fixed effects regression coefficients, the maximum likelihood estimators (MLE) are computed using an expectation conditional maximization either (ECME) algorithm which is variant of the expectation-maximization algorithm (McLachlan and Krishnan, 2007). We obtained MLE via the Laplace and quadrature estimation (Stroup, 2013).

The Inference about each of the fixed effects terms and covariance parameters is performed using Wald-type tests with the estimated variance-covariance matrix (Stroup, 2013). The

global effect of a covariate on the multiple outcomes is tested via a type-III F test. This test in Proc GLIMMIX is reported as the “Type III Tests of Fixed Effects table “ and contains hypothesis tests for the significance of each of the fixed effects or their combinations.

3.4. Fitting the mGLMM in SAS

In SAS Proc GLIMMIX (SAS 9.3, SAS Institute Inc.), fitting these models can be made as follows. First, since Proc GLIMMIX requires a univariate data structure, the 11 responses at each frequency are stacked as 11 observations per subject at each time point and a new variable that indicates the frequency (freq_type) is included to distinguish between the 11 responses. After arranging the data as in WebTable 2, the SAS macro provided in the Appendix can be used to implement these approaches. We used both pseudo-likelihood (Breslow and Clayton, 1993; Wolfinger, 1993; Wolfinger and O’Connell, 1993) and maximum likelihood via AGQ (Evans, 1993; Pinheiro and Chao, 2006) to estimate the parameters. It is important to note that the problem of estimation of model parameters becomes more difficult due to the infiltration of additional parameters from the covariance matrix of the random effect terms especially when analyzing large data sets.

The estimation can also be made within Proc MIXED (Gao Feng et al., 2003; Thiebaut Rodolphe et al., 2002). But, the capability of the codes suggested using Proc MIXED can be generalized via GLIMMIX. Moreover, there are some limitations of their MIXED based code such as (i) their codes are only studied under bivariate normal outcomes. Especially, the SAS application to the Kronecker product approach is limited to bivariate normal outcomes. The random coefficient approach was tested for 3 outcomes. We know that there are multiple issues of convergence that one faces when trying to fit such models for more than 3 outcomes. (ii) The covariance structure they use, AR(1), assumes repeated measures are equally spaced for all subjects which does not hold in many longitudinal biomedical studies. (iii) Their SAS program allows to estimate only one correlation parameter for the ‘bivariate process’ rather than a correlation matrix, which forces the practitioner to assume that the intra-marker correlation is the same for the two markers and the inter-marker correlation is proportional to the intra-marker correlation. (iv) when the measurements of the two markers never occur at the same time because of a design consideration, their SAS code based on AR process would fail to account for this type of missingness. Their Random coefficient SAS code will have difficulty of convergence to deal with more than a few biomarkers too. Our code, under the constraints we discussed, can deal with these problems seamlessly. (v) Finally, their code only works for Gaussian distributed outcomes and can’t be used with mixed outcomes (continuous, binary, count).

3.5. Model fit assessment and selection

The GLIMMIX procedure in SAS (SAS 9.3) computes various information criteria that typically apply a penalty to the log likelihood ($-2\log L$). We used the $-2\log L$, Akaike’s information criteria (AIC) (Akaike, 1974), and Schwarz’s Bayesian criterion (BIC) (Bozdogan, 1987) to make comparisons among the mGLMM variations. The smaller the value of these fit statistics, the better the model relative to its competitors. Finally, we assessed the properties and model assumptions (e.g., normality of residuals) of the best fitting model using residual analysis (see Web Figure 4).

3.6. Pseudo-simulation study

Since the goal here is to examine the standard error estimates from the empirical estimation techniques considered in the study rather than to examine the finite sample performance of the methods, we used a bootstrap approach. We used 500 repeated re-samplings (bootstrap samples) of the full data to assess and compare the methods discussed above and to assess the robustness of the variance estimators. It was not feasible to run more than 500 bootstrap samples as the time it took for models like hSPRI alone was up to 60 hours.

$$\begin{aligned}
 y_{i1} &= -0.834 + 0.745time + 0.383age + -8.938male + 0.155age * male \\
 &\quad (3.267) \quad (0.051) \quad (0.049) \quad (5.195) \quad (0.078) \\
 y_{i2} &= -4.917 + 0.946time + 0.553age + -7.616male + 0.240age * male \\
 &\quad (3.267) \quad (0.051) \quad (0.049) \quad (5.195) \quad (0.078) \\
 y_{i3} &= -7.933 + 0.930time + 0.667age + 4.317male + 0.228age * male \\
 &\quad (3.268) \quad (0.051) \quad (0.049) \quad (5.225) \quad (0.078) \\
 y_{i4} &= -9.299 + 0.894time + 0.777age + 7.368male + 0.228age * male \\
 &\quad (3.985) \quad (0.051) \quad (0.059) \quad (6.339) \quad (0.095) \\
 y_{i6} &= \sim 3.113 + 1.098time + 0.815age + 0.055male + 0.269age * male \\
 &\quad (3.995) \quad (0.051) \quad (0.060) \quad (6.340) \quad (0.095) \\
 y_{i8} &= -12.36 + 1.269time + 1.070age + 7.6538male + 0.100age * male \\
 &\quad (3.733) \quad (0.059) \quad (0.056) \quad (5.942) \quad (0.089) \\
 y_{i9} &= -9.896 + 1.153time + 1.203age + 6.534male + 0.083age * male \\
 &\quad (3.739) \quad (0.059) \quad (0.056) \quad (5.954) \quad (0.089) \\
 y_{i10} &= -5.558 + 1.194time + 1.266age + 2.0938male + 0.089age * male \\
 &\quad (3.742) \quad (0.060) \quad (0.056) \quad (5.960) \quad (0.089)
 \end{aligned} \tag{1.5}$$

This was implemented via a non-parametric bootstrapping approach (Efron and Tibshirani, 1993) with a repeated sample of the observed data with replacement with the study subject as the sampling unit to form 500 datasets of each with 882 sample size in 8 different frequencies (considering >8 outcomes was computationally demanding).

The real nature of the variability in these outcomes and the fact that we considered 8 outcomes makes this interesting. It is difficult to design hypothetical simulation studies that reflect what would have been observed in real studies like ours. The parameters in equation (1.5) are estimates from the original data set (fixed effect estimates and corresponding standard errors). Traditionally, Monte-Carlo simulation studies based on data generated from statistical models have been used for comparing different methods. However, it is not easy to simulate data that reflects reality when the number of outcome is as large as what we have in our data example. Resampling has the advantage that the data in resampled datasets are based on observations from real datasets and thus reflect the appropriate level of diversity and variability found in actual populations (Efron and Tibshirani, 1993; Marshall, Altman and Holder, 2010). Performance of methods was compared via AIC, BIC, $-2\log L$, and via

comparison of relative precision of parameter estimates. Of course, theoretical finite sample examination of these methods can be done using a small number of outcomes.

4. Results

The results of our analyses are reported in four main tables and three figures. Additional results are reported in the Web supplementary materials. Figure 1 depicts the mean trajectories over time of hearing loss outcomes for each frequency. It shows an increasing linear trend with rates slightly increasing for each higher frequency. This was consistent with the non-significant quadratic effects of time and age at entry. WebFigure 1 shows the profiles of individual trajectories at 3 kHz and 8 kHz and depicts the trajectories of hearing thresholds of older adults between right and left ears by gender. Hearing thresholds at other frequencies were similar. The figures show an increasing trend in hearing loss over time for each frequency with no significant differences in the variability of individual threshold measures between the two ears, which favors the nested effects model over the crossed effects model (Faraway, 2004). The nested approach doubles the number of observations used to estimate parameters while adding an extra parameter for the variance component of the random effect to account for nesting and clustering of observations by ear.

Table 1 shows the goodness of fit estimates based on likelihood information criteria comparing the different mGLMM approaches: SHRI, SHRIS, SPRI, SPRIS, hSPRI and hSPRIS. While SPRIS is the best in the simulated data, hSPRIS followed by HSPRI showed better performance in the real data. Given that the SPRIS is nested within hSPRIS with a difference of 3 parameters and the difference in $-2\log L$ of the two approaches, $244.3=652.7-408.4$, is significant ($p<0.05$) using a chi-square test with 3 degrees of freedom, we can conclude that hSPRIS is a superior approach to modeling multiple hearing loss outcomes measured at different frequencies.

The parameter estimates and their standard errors from the simulation study are reported in Figures 2. The six curves in each panel represent the distribution of the estimates (Beta or SE) for SHRI, SHRIS, SPRI, SPRIS, hSPRI and hSPRIS. Figure 2 (first row) shows that both SPRI, SPRIS, hSPRI and hSPRIS provide consistent estimates of all the parameters and Figure 2 (2nd row) shows that hSPRIS leads to the lowest standard error estimates. The asymptotic normality of the distribution of betas is met in SPRI, SPRIS, hSPRI and hSPRIS, but not in SHRI and SHRIS. Except for SHRI and SHRIS, all appear to give consistent estimates of the true value of beta_age (0.383 at 1 kHz, 0.553 at 2 kHz, etc.). Similarly, the SE or precision estimates of SHRI and SHRIS were much higher than those SPRI, SPRIS, hSPRI and hSPRIS, with SPRIS and hSPRIS showing the smallest SE.

Table 2 shows the fixed effects parameter estimates under the separate (sGLMM) and multivariate GLMM (mGLMM) for our data sample. The coefficients of time measure the longitudinal changes in thresholds for each individual after adjusting for age at entry, gender, and their interactions. The results appropriately model the expected increase in hearing loss over time for each frequency (except at 14kHz), and uniquely show the rate of hearing loss increasing with frequency until 10 kHz, where the rate of hearing loss change starts to

plateau and then decreases at 14 kHz even though the quadratic terms in our model did not show significance.

The coefficients for age measure the cross-sectional difference in threshold in individuals of different ages after adjusting for covariates. As noted earlier, hearing loss increases with age with the magnitude of hearing loss change increasing with frequency. For example, a subject who was 10 years older at baseline has a 3.8 dB increase in hearing loss at 1 kHz, 5.5 dB increase at 2 kHz, 12.6 dB increase at 11 kHz and 12.9 dB increase at 12 kHz. However, after adjusting for age at entry, the rate of change over three years of observation at 1 kHz, 3 kHz, 8 kHz, and 12 kHz were 0.75 dB, 0.93 dB, 1.27 dB, and 0.67 dB respectively.

There were also significant gender differences at each frequency with males having less hearing loss than females at 1 kHz and 2 kHz, but more hearing loss at 3 kHz and higher frequencies (consistent with conclusions from animal models of metabolic and sensory pathologies underlying age-related hearing loss (Dubno et al., 2013; Schmiedt et al., 2002)). Fixed effects estimates from mGLMM and sGLMM are very similar for the data example.

Table 3 shows the asymptotic standard error estimates for the fixed effect terms comparing sGLMM to mGLMM. The standard error estimates from mGLMM are, in most cases, slightly less than those from sGLMM, indicating more accurate estimates of precision of the fixed effect parameters. A possible explanation is that the former accounts for the correlations among the multiple outcomes and dependency due to clustering by ear. It has been noted that ignoring this dependency could lead to incorrect inference due to inflation of type-I error (Berridge and Crouchley, 2011; Gibbons, Hedeker and DuToit, 2010).

Table 4 shows the Type III F-test for the global hypothesis of covariate effect on the multiple outcomes using the six different mGLMM approaches. All methods lead to the same general conclusion, with significant time, age at entry, gender, and age and gender interaction effects. These global tests indicate that the effect of covariates across frequency thresholds is different. The level of significance varied among approaches with hSPRIS resulting in the most conservative results for nearly all of the covariate terms in the model, followed by hSPRI. Additional figures for attrition rates over time (Web Figure2), trajectory of mean thresholds over time (WebFigure 3), and model diagnostic for normality and correlation of responses (WebFigure4) are reported in the supplementary materials.

Based on these results, the two best approaches for modeling multiple thresholds measured longitudinally are hSPRIS (that assumes heterogeneity in the measurement error variance among a group of frequencies) and SPRIS (that assumes homogeneity in the measurement error variance among individual test frequencies). Both gave lower AIC/BIC estimates

5. Discussion

The most common analysis of pure-tone thresholds involves modeling of the outcomes at each frequency for each ear using separate generalized linear mixed models (sGLMM). This type of analysis makes sub-optimal use of extensive data on age-related hearing loss, ignores crucial information in the data, may fundamentally underestimate or overestimate the magnitude and significance of the association between hearing loss and covariates and may

not answer key questions such as, (1) testing the joint effect of covariates on multiple outcomes, (2) estimating the joint evolution of the multiple outcomes over time and (3) estimating the association between the evolutions of the multiple outcomes. Thus, the key contributions of this new analysis are: (1) a unified framework for the analysis of multivariate longitudinal hearing thresholds based on mGLMM; (2) an optimal approach to dealing with dependency of responses from two ears due to nesting of effects; (3) a global test for covariate effect on multiple outcomes; and (4) a SAS macro to implement the proposed approach.

The proposed mGLMM approach has the ability to incorporate dependency in the threshold measures within and across individuals that arise due to the repeated measures over time, nesting and clustering by ear, multiple frequencies and random error. This approach provides a more efficient estimation of the parameters in the model (Shah, Laird and Schoenfeld, 1997), understanding the interdependence across variables and patterns of change in the thresholds (Gebregziabher et al., 2013; MacCallum et al., 1997). Thus, given that individual variation is a key feature of age-related hearing loss (Lee et al., 2005; Moller, 2006), our results show that mGLMMs are suitable and superior in accurately modeling such variations compared to sGLMM. Compared to sGLMM, mGLMMs provides more accurate estimates of the standard errors (see table 3) reducing the likelihood of committing type-I error that is higher than the nominal level. In general, the mGLMM approach provides a unified framework for analyzing multiple longitudinal outcomes allowing for trade-off among model parsimony, precision, and computational burden.

Although hearing loss at mid-to-higher frequencies is expected to increase with increasing frequency and with increasing age and over time (e.g. (Cruickshanks et al., 1998)), this was not the case in the current dataset that included extended higher frequencies (8 kHz). A recent study (Davidov and Rosen, 2011) has proposed a restricted maximum likelihood approach wherein parameters are restricted to maintain the ordering of effects by age, over time, and over increasing frequency. They imposed a set of constraints on the frequency-specific fixed effect parameters and used constrained expected-conditional maximization (CECME) to estimate them. Although this approach leads to parameter estimates that are naturally ordered, it also requires a CECME algorithm that may not be easily accessible in standard software packages. In contrast, our approach does not provide fully constrained and ordered estimates, consistent with the data, but leads to estimates that do not highly violate the ordering.

A common constraint in the analysis of longitudinal data is the presence of missing data. Web Figure 3 depicts the increasing degree of missingness with increasing frequency, which has the potential for introducing bias in the parameter estimates. If the missingness occurs at random, maximum likelihood or restricted likelihood using available cases are known to lead to consistent parameter estimates (Little and Rubin, 1987). In our example, the mechanisms for missing thresholds include the following classifications: (i) missing data due to right censoring (for example, as a result of damage to the auditory system) where thresholds are measured longitudinally for all time points up to the censoring frequency but are missing for the remaining frequencies, and/or data are incomplete due to missing one or more visits in which the thresholds for all measures that would have been obtained in those

visits are unobserved. The first mechanism could lead to missingness that cannot be ignored in a separate analysis of each threshold; nevertheless, a joint modeling approach provides valid inference because thresholds at different frequencies are not independent and extra information can be used from the opposite ear as both ears may not be censored at the same frequency, making the missingness ignorable. The other challenge with fitting mGLMM models is the problem of convergence and the time for these models to converge. For example, in our simulation study of 11 outcomes at six time points and sample size 882, SHRI and SHRIS took about 15 minutes each and SPRI took 330 minutes, while SPRIS took 60 hours (for 500 simulations). Both hSPRI and hSPRIS took over 60 hours each.

Our results show that mGLMM is promising and hence further investigations to understand the operating characteristics under small sample size via simulations studies are needed. We focused our mGLMM analysis on multiple Gaussian outcomes with Gaussian random effects, but the extension to non-Gaussian outcomes and non-Gaussian random effects could be made in a straightforward fashion as is done in the GLMM framework.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix

Sample SAS code for implementing the joint modeling approaches discussed in the paper. After arranging the data as in WebTable 2 of the paper, then the following SAS codes can be used to implement the variations of mGLMM.

Note: fr_type2 is some grouped form of fr_type based on the variability profiles of HL outcomes in each frequency.

```
%macro mGLMM1(meth, distn, randm1, randm2);
```

```
proc glimmix data=fulldata;
```

```
class fr_type fr_type2;
```

```
model hl=fr_type fr_type*timec fr_type*entryage fr_type*male fr_type*male*entryage/  
dist=&distn s noint DDFM=RESIDUAL CL ;
```

```
random &randm1 &randm2 / subject=subject type=chol g gcorr;
```

```
random lear/ subject=subject; *accounts for nesting by ear;
```

```
random _residual_ / subject=subject type=VC; *accounts for R-side;
```

```
%mend mGLMM1;
```

```
%macro mGLMM2(meth, distn, randm1, randm2);
```

```
proc glimmix data=fulldata ;
```

```

class fr_type fr_type2;

model hl=fr_type fr_type*timec fr_type*entryage fr_type*male fr_type*male*entryage/
dist=&distn s noint DDFM=RESIDUAL CL ;

random &randm1 &randm2/ subject=subject type=chol group=fr_type2 g gcorr;

random lear/ subject=subject group=fr_type2;

random _residual_ / subject=subject type=VC; *no group statement here;

%mend mGLMM2;

%macro mGLMM3(meth, distn, randm1, randm2);

proc glimmix data=fulldata ;

class fr_type fr_type2;

model hl=fr_type fr_type*timec fr_type*entryage fr_type*male fr_type*male*entryage /
dist=&distn s noint DDFM=RESIDUAL CL ;

random &randm1 &randm2/ subject=subject type=chol group=fr_type2 g gcorr;

random lear/ subject=subject group=fr_type2; run;

random _residual_ / subject=subject type=VC group=fr_type2;

%mend mGLMM3;

*****

*ANALYSIS

*****;

title "SHRI";

%mGLMM1(SHRI,normal, int, ); * SHRI;

title "SHRIS";

%mGLMM1(SHRIS, normal, int, timec); *SHRIS;

title "SPRI common R-side variance";

%mGLMM2(SPRI, normal, int, ); * SPRI;

title "SPRIS common R-side variance";

%mGLMM2(SPRIS, normal, int, timec); *SPRIS;

```


title “SPRI heterogeneous R-side variance- replacing fr_type2 by fr_type leads to analysis equivalent to separate analysis”;

```
%mGLMM3(HSPRI, normal, int, ); * HSPRI;
```

title “SPRIS heterogeneous R-side variance - replacing fr_type2 by fr_type leads to analysis equivalent to separate analysis”;

```
%mGLMM3(HSPRIS, normal, int, timec); *HSPRIS;
```

References

- Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Berridge DM, and Crouchley R (2011). *Multivariate generalized linear mixed models using R* Boca Raton, FL: CRC Press.
- Bozdogan H (1987). Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52, 345–370.
- Breslow NE, and Clayton DG (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 9–25.
- Cruikshanks KJ, Wiley TL, Tweed TS, et al. (1998). Prevalence of Hearing Loss in Older Adults in Beaver Dam, Wisconsin: The Epidemiology of Hearing Loss Study. *American Journal of Epidemiology* 148, 879–886. [PubMed: 9801018]
- Davidov O, and Rosen S (2011). Constrained inference in mixed-effects models for longitudinal data with application to hearing loss. *Biostatistics* 12, 327–340. [PubMed: 20719782]
- Diggle P, Heagerty P, Liang K-Y, and Zeger S (2002). *Analysis of longitudinal data* New York, NY: Oxford University Press.
- Dubno J, Eckert M, Lee F-S, Matthews L, and Schmiedt R (2013). Classifying Human Audiometric Phenotypes of Age-Related Hearing Loss from Animal Models. *Journal of the Association for Research in Otolaryngology* 14, 687–701. [PubMed: 23740184]
- Dullemeijer C, Verhoeef P, Brouwer IA, Kok FJ, Brummer RJM, and Durga J (2010). Plasma very long-chain N-3 polyunsaturated fatty acids and age-related hearing loss in older adults. *The journal of nutrition, health & aging* 14, 347–351.
- Echt KV, Smith SL, Burridge AB, and Spiro A (2010). Longitudinal changes in hearing sensitivity among men: The Veterans Affairs Normative Aging Study. *The Journal of the Acoustical Society of America* 128, 1992–2002. [PubMed: 20968370]
- Efron B, and Tibshirani RJ (1993). *An introduction to the bootstrap*, 1993 Chapman & Hall, New York.
- Evans G (1993). *Practical numerical integration*: Wiley New York.
- Faraway JJ (2004). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*: CRC press.
- Fieuws S, and Verbeke G (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Stat Med* 23, 3093–3104. [PubMed: 15449333]
- Fieuws S, and Verbeke G (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 62, 424–431. [PubMed: 16918906]
- Fieuws S, Verbeke G, and Molenberghs G (2007). Random-effects models for multivariate repeated measures. *Stat Methods Med Res* 16, 387–397. [PubMed: 17656450]
- Fitzmaurice GM, Laird NM, and Ware JH (2004). *Applied longitudinal analysis*
- Feng Gao, Paul Thompson, Chengjie Xiong, and Philip M (2003). *Analyzing Multivariate Longitudinal Data Using SAS* Cary, NC: SAS Institute Inc

- Gebregziabher M, Zhao Y, Dismuke CE, Axon N, Hunt KJ, and Egede LE (2013). Joint modeling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models. *Health Services and Outcomes Research Methodology* 13, 39–57. [PubMed: 30555270]
- Gibbons RD, Hedeker D, and DuToit S (2010). Advances in Analysis of Longitudinal Data. *Annual Review of Clinical Psychology* 6, 79–107.
- Kiely KM, Gopinath B, Mitchell P, Luszcz M, and Anstey KJ (2012). Cognitive, Health, and Sociodemographic Predictors of Longitudinal Decline in Hearing Acuity Among Older Adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 67, 997–1003.
- Lee F-S, Matthews LJ, Dubno JR, and Mills JH (2005). Longitudinal study of pure-tone thresholds in older persons. *Ear and hearing* 26, 1–11. [PubMed: 15692300]
- Little R, and Rubin D (1987). *Statistical analysis with missing data* New York: John Wiley & Sons.
- Liu L, Conaway MR, Knaus WA, and Bergin JD (2008). A random effects four-part model, with application to correlated medical costs. *Computational Statistics & Data Analysis* 52, 4458–4473.
- MacCallum RC, Kim C, Malarkey WB, and Kiecolt-Glaser JK (1997). Studying Multivariate Change Using Multilevel Models and Latent Curve Models. *Multivariate Behavioral Research* 32, 215–253. [PubMed: 26761610]
- Marshall A, Altman DG, and Holder RL (2010). Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC medical research methodology* 10, 112. [PubMed: 21194416]
- McLachlan G, and Krishnan T (2007). *The EM algorithm and extensions*: John Wiley & Sons.
- Moller AR (2006). *Hearing: anatomy, physiology, and disorders of the auditory system*: Academic press.
- Morrell CH, and Brant LJ (1991). Modelling hearing thresholds in the elderly. *Statistics in Medicine* 10, 1453–1464. [PubMed: 1925173]
- Park SK, Elmarsafawy S, Mukherjee B, et al. (2010). Cumulative lead exposure and age-related hearing loss: The VA Normative Aging Study. *Hearing Research* 269, 48–55. [PubMed: 20638461]
- Pinheiro JC, and Chao EC (2006). Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics* 15, 58–81.
- Rosen S, and Davidov O (2012). Order-restricted inference for multivariate longitudinal data with applications to the natural history of hearing loss. *Stat Med* 31, 1761–1773. [PubMed: 22729892]
- Schmiedt RA, Lang H, Okamura H. o., and Schulte BA (2002). Effects of furosemide applied chronically to the round window: a model of metabolic presbycusis. *The Journal of neuroscience* 22, 9643–9650. [PubMed: 12417690]
- Shah A, Laird N, and Schoenfeld D (1997). A Random-Effects Model for Multiple Characteristics with Possibly Missing Data. *Journal of the American Statistical Association* 92, 775–779.
- Stroup W (2013). *Generalized Linear Mixed Models* Boca Raton, FL: CRC Press.
- Thiebaut Rodolphe, Jacqmin-Gadda Helene, Genevieve C, Catherine L, and Daniel C (2002). Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine* 69 249–256. [PubMed: 12204452]
- Verbeke G, Fieuws S, Molenberghs G, and Davidian M (2014). The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res* 23, 42–59. [PubMed: 22523185]
- Verbeke G, Spiessens B, and Lesaffre E (2001). Conditional Linear Mixed Models. *The American Statistician* 55, 25–34.
- Wiley TL, Chappell R, Carmichael L, Nondahl DM, and Cruickshanks KJ (2008). Changes in hearing thresholds over 10 years in older adults. *J Am Acad Audiol* 19, 281–292; quiz 371. [PubMed: 18795468]
- Wolfinger R (1993). Laplace's Approximation for Nonlinear Mixed Models. *Biometrika* 80, 791–795.
- Wolfinger R, and O'Connell M (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48, 233–243.

Figure 1:
Trajectory of mean thresholds over time for each frequency (for left ear, the comparison with right ear is given in WebFigure 4)

Figure 2:

Distribution of Mean Beta (Beta) and Standard error (SE) estimates of selected variables in mGLMM model (first row: for age, second row: for age*male) for each method at a selected frequency (1 kHz). Based on 500 bootstrap samples.

Table 1:

Statistical Information Statistics to compare joint models

	Simulation			Real Data		
Method	-2logL	AIC	BIC	-2logL	AIC	BIC
SHRI	196250.9	196256.9	196269.9	242556.4	242562.4	242576.8
SHRIS	196099.1	196108.8	196129.7	242490.1	242500.1	242524.0
SPRI	185267.9	185285.9	185324.9	232326.3	232344.3	232387.4
SPRIS	184193.7	184223.1	184286.5	231652.7	231686.7	231768.0
hSPRI	185514.6	185528.6	185558.9	232168.6	232192.6	232250.6
hSPRIS	184616.9	184641.8	184695.5	231408.4	231448.4	231544.1

SHRI-shared random intercept; SHRIS- shared random intercept and slope; SPRI-separate random intercept; SPRIS- separate random intercept and slope; hSPRI-heterogenous separate random intercept; hSPRIS- heterogenous separate random intercept and slope

AIC-Akaike information criterion, BIC- Bayesian information criterion

Table 2:

Fixed effect parameter estimates comparing separate (sGLMM) versus joint (mGLMM) under the random intercept and slope (RIS) framework for real data

Method	Parameter	Intercept	Time (years)	Age	Male	Age*Male
sGLMM (RIS)	01k	-0.835	0.745	0.383	-8.938	0.155
	02k	-4.918	0.946	0.553	-7.616	0.239
	03k	-7.934	0.930	0.667	4.317	0.228
	04k	-9.294	0.894	0.777	7.360	0.229
	06k	3.119	1.098	0.814	0.046	0.269
	08k	-12.35	1.269	1.070	7.654	0.096
	09k	-9.893	1.153	1.203	6.534	0.083
	10k	-5.555	1.194	1.266	2.093	0.089
	11k	-1.659	0.869	1.259	6.790	-0.022
	12k	4.821	0.666	1.291	6.502	-0.028
	14k	25.96	-0.034	1.128	7.618	-0.064
mGLMM (SPRIS)	01k	-0.799	0.747	0.383	-8.977	0.155
	02k	-4.888	0.948	0.553	-7.650	0.239
	03k	-7.889	0.932	0.666	4.273	0.227
	04k	-9.390	0.891	0.779	7.515	0.226
	06k	3.017	1.095	0.817	0.205	0.226
	08k	-12.36	1.265	1.071	7.683	0.095
	09k	-9.893	1.150	1.203	6.592	0.082
	10k	-5.545	1.192	1.266	2.149	0.088
	11k	-1.608	0.864	1.259	6.771	-0.021
	12k	4.874	0.660	1.290	6.483	-0.028
	14k	25.99	-0.037	1.127	7.604	-0.064

RIS- random intercept and slope; GLMM- generalized linear mixed model, age at entry

Table 3:

Standard Error estimates of the fixed effects comparing separate (sGLMM) versus joint (mGLMM) under the random intercept and slope (RIS) framework for real data

Method	Parameter	Intercept	Time	Age	Male	Age*Male
sGLMM (RIS)	01k	3.268	0.0505	0.0491	5.197	0.078
	02k	3.268	0.0505	0.0491	5.197	0.078
	03k	3.284	0.0517	0.0493	5.226	0.078
	04k	3.987	0.0507	0.0599	6.341	0.095
	06k	3.996	0.0520	0.0600	6.362	0.095
	08k	3.735	0.0585	0.0560	5.945	0.089
	09k	3.740	0.0594	0.0561	5.957	0.089
	10k	3.744	0.0601	0.0562	5.963	0.089
	11k	2.966	0.0662	0.0450	4.734	0.072
	12k	2.993	0.0715	0.0456	4.786	0.073
	14k	3.042	0.1023	0.0470	4.909	0.077
mGLMM (SPRIS)	01k	3.254	0.0484	0.0488	5.176	0.078
	02k	3.254	0.0484	0.0489	5.176	0.078
	03k	3.267	0.0494	0.0491	5.201	0.078
	04k	3.977	0.0517	0.0597	6.324	0.095
	06k	3.987	0.0530	0.0599	6.348	0.095
	08k	3.732	0.0586	0.0560	5.940	0.089
	09k	3.738	0.0596	0.0561	5.953	0.089
	10k	3.742	0.0601	0.0562	5.959	0.089
	11k	2.968	0.0660	0.0450	4.737	0.072
	12k	2.996	0.0716	0.0457	4.792	0.073
	14k	3.049	0.1038	0.0471	4.922	0.077

RIS- random intercept and slope; GLMM- generalized linear mixed model, age at entry

Table 4:

Type III F test and p-values for testing global covariate effect or whether each of the covariate effects (time, age, male, age*male) are equal for all frequencies

	Parameter	Intercept	Time	Age	Male	Age*Male
SHRI	F	27.6	193.3	195.8	4.3	2.3
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.008
SHRIS	F	27.7	115..7	196.9	4.3	2.3
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.008
SPRI	F	32.5	359.6	176.9	4.5	2.3
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.008
SPRIS	F	34.0	161.0	169.3	4.8	2.2
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.011
hSPRI	F	34.0	372.9	175.0	4.3	2.3
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.010
hSPRIS	F	36.2	159.6	167.5	4.5	2.2
	P value	<0.0001	<0.0001	<0.0001	<0.0001	0.014

SHRI-shared random intercept; SHRIS- shared random intercept and slope; SPRI-separate random intercept; SPRIS- separate random intercept and slope; hSPRI-heterogenous separate random intercept; hSPRIS- heterogenous separate random intercept and slope, age=age at entry