FEMA-Long: Modeling unstructured covariances for discovery of time-dependent effects in large-scale longitudinal datasets

List of authors:

Pravesh Parekh[1,*], Nadine Parker[1], Diliana Pecheva[3], Evgeniia Frei[1], Marc Vaudel[4,5,6], Diana M. Smith[7,8], Alison Rigby[9], Piotr Jahołkowski[1], Ida Elken Sønderby[1,10], Viktoria Birkenæs[1,11], Nora Refsum Bakken[1], Chun Chieh Fan[12], Carolina Makowski[13], Jakub Kopal[1], Robert Loughnan[3,14], Donald J. Hagler Jr.[3], Dennis van der Meer[1], Stefan Johansson[4,15], Pål Rasmus Njølstad[4,15], Terry L. Jernigan[2,7,13,16], Wesley K. Thompson[12], Oleksandr Frei[1,17], Alexey A. Shadrin[1], Thomas E. Nichols[18,19], Ole A. Andreassen[1], Anders M. Dale[2,3,13,16,20,*]

Affiliations:

[1]Centre for Precision Psychiatry, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, Oslo, Norway

[2]Department of Radiology, School of Medicine, University of California San Diego, La Jolla, CA, USA

[3]Center for Multimodal Imaging and Genetics, J. Craig Venter Institute, La Jolla, CA, USA

[4]Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway

[5]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

[6]Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Bergen, Norway

[7]Center for Human Development, University of California San Diego, La Jolla, CA, USA

[8]Medical Scientist Training Program, University of California San Diego, La Jolla, CA, USA

[9]Neuroscience Graduate Program, University of California San Diego, La Jolla, CA, USA

[10]Department of Medical Genetics, Oslo University Hospital & University of Oslo, Oslo, Norway

[11]PsychGen Center for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway

[12]Center for Population Neuroscience and Genetics, Laureate Institute for Brain Research, Tulsa, OK, USA

[13]Department of Psychiatry, School of Medicine, University of California San Diego, La Jolla, CA, USA

[14]Population Neuroscience and Genetics Lab, University of California San Diego, La Jolla, CA, USA

[15]Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen, Norway

[16]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA

[17]Department of Pharmacy, Section for Pharmacology and Pharmaceutical Biosciences, University of Oslo, Oslo, Norway

[18]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

[19]Oxford Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[20]Department of Neuroscience, University of California San Diego, La Jolla, CA, USA

*Correspondence should be addressed to:

Dr. Pravesh Parekh,
Centre for Precision Psychiatry,
Ullevål Hospital, Building 48,
Oslo 0450, Norway
praveshparekh@gmail.com

Prof. Anders M. Dale,
J. Craig Venter Institute,
4120 Capricorn Lane,
La Jolla, CA 92037, USA,
andersmdale@gmail.com

Abstract

Linear mixed-effects (LME) models are commonly used for analyzing longitudinal data. However, most applications of LME models rely on random intercepts or simple, e.g., stationary, covariance. Here, we extend the Fast and Efficient Mixed-Effects Algorithm (FEMA) and present FEMA-Long, a computationally tractable approach to flexibly modeling longitudinal covariance suitable for high-dimensional data. FEMA-Long can i) model unstructured covariance, ii) model non-linear fixed effects using splines, iii) discover time-dependent fixed effects with spline interactions, and iv) perform genome-wide association studies (GWAS) supporting discovery of time-dependent genetic effects. We applied FEMA-Long to perform a longitudinal GWAS with non-linear SNP-by-time interaction on length, weight, and body mass index of 68,273 infants with up to six measurements in the first year of life. We found dynamic patterns of random effects including time-varying heritability and correlations, as well as several genetic variants showing time-dependent effects, highlighting the applicability of FEMA-Long to enable novel discoveries.

Introduction

Large-scale longitudinal datasets are critical for novel discoveries in health and medicine and can transform healthcare[1]. Longitudinal data allows studying the dynamic processes underlying health and disease development/progression by modeling trajectories, examining time-dependent effects of variables, investigating association with health indicators and outcomes, and making predictions of outcomes. Examples of large-scale densely-sampled datasets include the UK Biobank (https://www.ukbiobank.ac.uk), All of Us (https://allofus.nih.gov), the Adolescent Brain Cognitive Development℠ Study (https://abcdstudy.org), and many others[2].

To leverage the full potential of these studies, we need computationally efficient modeling strategies, optimized for large-scale high-dimensional datasets. These strategies need to account for non-independence of observations (e.g., repeated measurements and relationships between individuals), other sources of variance (e.g., shared environment), covariance in the data (e.g., correlations across time), flexible modeling of longitudinal trajectories, and enabling discovery of time-dependent effects. Linear mixed-effects (LME) models are a powerful class of analytical strategies that can account for these issues in longitudinal data such as non-independence of the samples and other known sources of variance modeled as random effects, in addition to including variables with non-linear effects on the outcome and non-linear interactions. Examples of these random effects include the similarity of measurements for the same subject, variance explained by genetics (heritability), variance attributable to shared environment, etc.

When using LME models on longitudinal datasets, random effects should be specified as time-varying. For example, heritability can vary significantly over time[3–8]. Similarly, correlations of measurements within subjects can change over time (e.g., autocorrelation). However, time-varying random effects are seldom modeled, given the significant computational burden of fitting such models. Thus, researchers use simplified LME models like compound symmetry (constant covariance over time) or making assumptions about the form of covariance (e.g., autoregressive models) which may not fit the data well. In addition, there is a lack of LME modeling software that scales to the needs of high-dimensional data[9] such as when estimating the marginal effects of several million single nucleotide polymorphisms (SNPs) in genome-wide association studies (GWAS), or when fitting mass-univariate models to millions of outcome variables, as is common in neuroimaging. This has severely limited our ability to maximize the use of high-dimensional longitudinal datasets.

Previously, we developed the Fast and Efficient Mixed-Effects Algorithm (FEMA)[10] for fitting a large number of LME models in a matter of seconds to minutes. FEMA massively accelerates LME fitting by using a covariance-regression approach to estimate random effects parameters and then quantizes these for

vectorized fitting of many phenotypes simultaneously. For example, FEMA can fit ~169,000 LME models with ~10,000 observations in ~54 seconds, a speedup of more than 3000 times, compared to standard LME solvers. In the present work, we introduce FEMA-Long, a framework adapted and optimized for high-dimensional longitudinal datasets. The features of FEMA-Long include i) modeling time-varying random effects using unstructured covariance, ii) modeling non-linear effects using splines, iii) examining time-dependent effects using spline interactions, and iv) performing GWAS, supporting the discovery of time-dependent effects. Through extensive simulations we show that FEMA-Long accurately models known sources of variance in a computationally fast manner while maintaining false positive rates. We demonstrate the use of FEMA-Long by performing a longitudinal GWAS on length, weight, and body mass index (BMI) of 68,273 infants with up to six repeated measurements (299,442 observations) during the first year of life, a highly dynamic period of development, and discovering SNPs showing time-dependent effects.

Results

FEMA-Long framework

FEMA is a method for fitting LMEs for large samples and a large number of outcome variables in a computationally fast manner[10]. Here, we introduce FEMA-Long, a framework optimized for high-dimensional longitudinal datasets. Given a set of continuous phenotypes, fixed effects, and random effects (which, besides grouping variables, can also include the genetic relatedness matrix (GRM)), we calculate the ordinary least squares residuals of the phenotypes. Then, using a method of moments (MoM) estimator, we decompose the residual variance into variances explained by the random effects. FEMA-Long supports estimating unstructured covariance, where for every pair of timepoints, we use the MoM estimation, with and without non-negativity constraint, to estimate variance and covariance respectively. Finally, we use the generalized least squares (GLS) estimation to update the fixed effects coefficients. FEMA-Long also supports modeling phenotypes as smooth functions of continuous variables (e.g., age), typically a spline basis function, which can also be used for interactions. To scale FEMA for performing GWAS-like analyses, we use a two-step procedure: first, fit a base model without SNPs (stage-1); next, regress the base-model residuals on the residualized SNPs (for the fixed effects in the base model), resulting in the estimation of the marginal effect of each SNP on the phenotype. We also use this two-step procedure to model the interaction of each SNP with the basis functions to discover time-dependent genetic effects on the phenotype.

In the next sections, we present the results from the evaluation of our approach. All simulations assume two nested random effects: family and subject (variances attributed to shared environment and repeated measurements respectively), in addition to fixed effects and noise. First, we show the accuracy of the estimated parameters compared to the ground truth; second, we present the false positive rates; third, we

demonstrate the computational efficiency of FEMA-Long; finally, we show the application of FEMA-Long by performing a longitudinal GWAS on length, weight, and BMI in infants during the first year of life.

FEMA-Long models unstructured covariances

To evaluate the accuracy of FEMA-Long for modeling unstructured covariances, we created 84 simulations with up to five repeated measurements and differing amounts of variances (0.1 to 0.9) explained by fixed effects, time-varying family- and subject-level covariances, and noise, with the total phenotypic variance of 1. We repeated each simulation 50 times and for each iteration calculated the root mean squared error (RMSE) between the estimated parameters and the ground truth. For comparison, we also fitted the models using the glmmTMB[11] toolbox in R[12], which can model unstructured covariances.

For the fixed effects, the average (over 50 iterations) RMSE across 84 simulations was $0.0121 \pm 0.0018\ [0.0073 - 0.0150], r_{\text{true,FEMA}} = 0.996$ (**Figure 1a**, **Supplementary Figure S1**). For the random effects parameters, the average RMSE for the family effect was $0.0160 \pm 0.0058\ [0.0064 - 0.0312], r_{\text{true,FEMA}} = 0.997$ (**Figure 1b**, **Supplementary Figure S2**), and for the subject effect was $0.0122 \pm 0.0036\ [0.0051 - 0.0218], r_{\text{true,FEMA}} = 0.999$ (**Figure 1c**, **Supplementary Figure S3**), indicating that the estimates were close to the ground truth.

The RMSE between the estimates from FEMA-Long and from glmmTMB toolbox for the fixed effects was $0.0004 \pm 0.0001\ [0.0002 - 0.0007], r_{\text{glmmTMB,FEMA}} = 1.00$ (**Supplementary Figure S4**), for the family effect was $0.0059 \pm 0.0021\ [0.0022 - 0.0111], r_{\text{glmmTMB,FEMA}} = 0.9996$ (**Supplementary Figure S5**), and for the subject effect was $0.0058 \pm 0.0017\ [0.0023 - 0.0107], r_{\text{glmmTMB,FEMA}} = 0.9998$ (**Supplementary Figure S6**), showing that the FEMA-Long estimates were almost identical to glmmTMB.

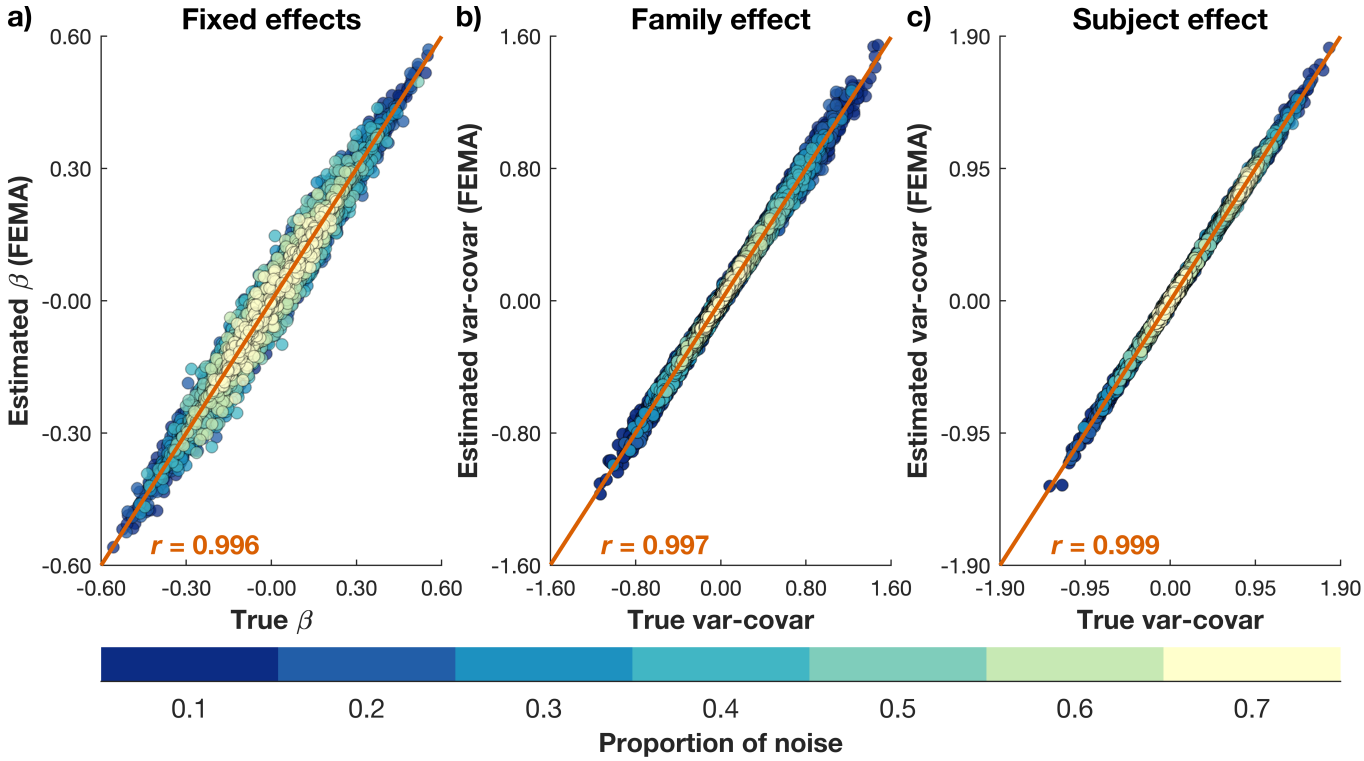

Figure 1: Comparison of estimated parameters from FEMA with ground truth. Each panel shows the scatterplot of the simulated ground truth and the estimated parameters from FEMA across 50 iterations of 84 simulation settings with different amounts of variances explained by fixed, family, and subject effects, as well as noise; the amount of noise in the simulated phenotype is indicated by the color of the points; the orange line indicates the least square fit and the correlation between ground truth and estimated parameters are indicated in orange text; **a)** beta coefficients for the fixed effects; **b)** variance-covariance components for the family random effect; and **c)** variance-covariance components for the subject effect.

FEMA-Long controls for false positives

To assess the number of false positives under the null hypothesis, we created 36 simulations with varying contribution from unstructured random effects and noise but no contribution from the fixed effects. For each simulation, we ran 1,000 iterations, with 100 fixed effects and 10 phenotypes. Then, we examined the distribution of $1000 \times 100 \times 10 - \log_{10} p$-values using unstructured and compound symmetry covariance patterns.

Across the 36 simulations, when using unstructured covariance, the p-values were uniformly distributed under the null ($-\log_{10} p$-values following the diagonal; green points in **Figure 2** and **Supplementary Figure S7**). However, for the compound symmetry covariance pattern, in almost all scenarios, the p-values had a non-uniform distribution with several of the p-values being smaller than expected (leftward deflection of the $-\log_{10} p$-values from the diagonal; orange points in **Figure 2** and **Supplementary Figure S7)**. These results show that the unstructured covariances p-values are well-calibrated.

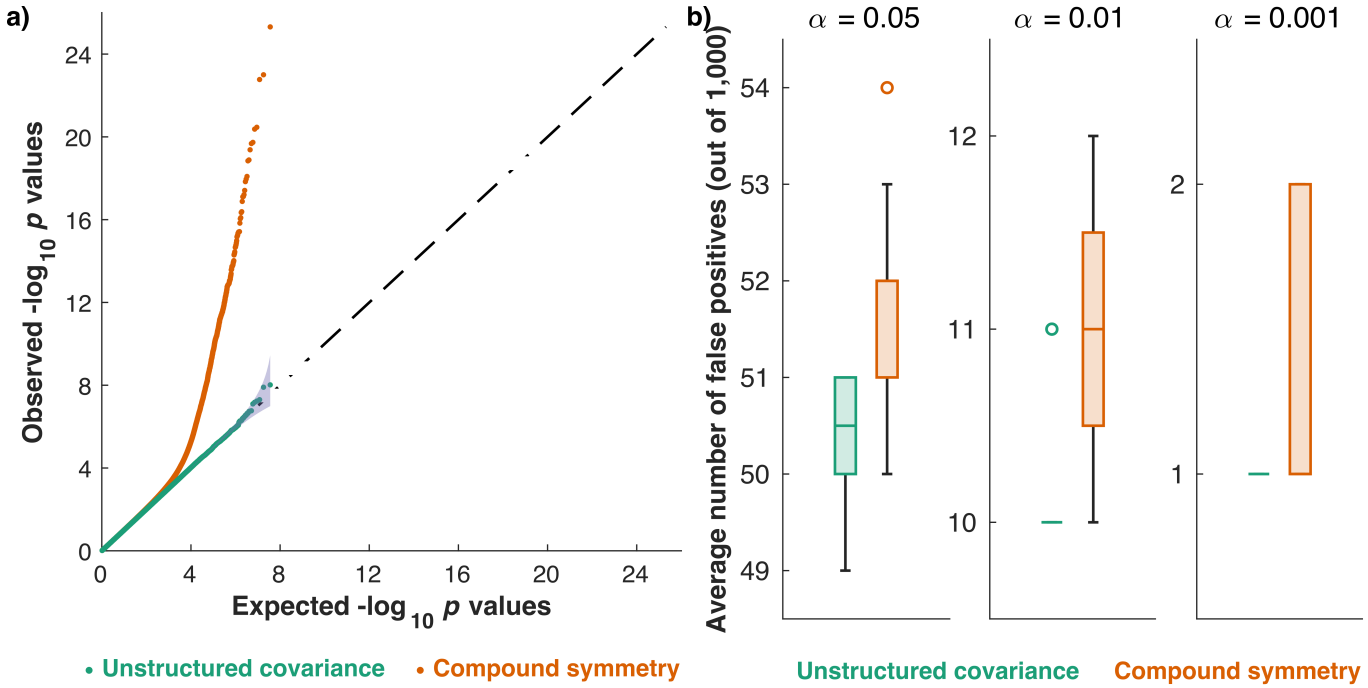

Figure 2: Examining false positives under the null. a) Distribution of $-\log_{10} p$-values across 1000 iterations of 36 simulation settings with differing amounts of variances explained by family effect, subject effect, and noise for unstructured covariance (green points) and compound symmetry (orange points); each simulation iteration consisted of 100 X variables and 10 outcome variables; the purple filled area indicates the 95% confidence interval based on inverse beta distribution; **b)** box-plots showing the average number of false positives (rounded off) at different alpha values across 100 X variables and 10 y variables for unstructured covariance (green bars and outlier dots) and compound symmetry (orange bars and outlier dots); the whiskers mark the values 1.5 times the interquartile range away from 75th and 25th percentile, and the dots mark the values which fall outside this range; under the null, the expected number of false positives are 50 ($\alpha = 0.05$), 10 ($\alpha = 0.01$), and 1 ($\alpha = 0.001$).

FEMA-Long is computationally efficient

We examined the computation time for FEMA-Long, with serial and parallel computing (32 parallel processes), for 50 fixed effects, varying the number of observations ($1{,}000 - 100{,}000$) and number of phenotypes ($1 - 10{,}000$). We also estimated the time taken by glmmTMB for a single phenotype.

FEMA-Long was fast across a range of observations and outcomes (**Figure 3**). For a single outcome, for 1,000 observations, FEMA took ~0.14 seconds while glmmTMB took ~9.95 seconds (i.e., FEMA-Long was ~71 times faster); for 100,000 observations, this time went up to ~37.27 seconds (FEMA-Long; **Figure 3a**) and ~1032.75 seconds (glmmTMB; **Supplementary Table S1**); therefore, for a single phenotype, FEMA-Long was between 27-121 times faster than glmmTMB.

For 100,000 observations and 10,000 outcome variables, FEMA-Long took ~63 minutes (serial, **Figure 3b**) and ~6.7 minutes (parallel, **Figure 3c**); in contrast, extrapolating the timing for glmmTMB, the same analysis would have taken ~119.5 days. Overall, for more than one phenotype, across a range of observations, FEMA was between ~85-2,893 times faster than glmmTMB (serial) and between ~81-29,620 times faster than glmmTMB (parallel; **Supplementary Table S2**). We encountered convergence issues with glmmTMB; increasing the maximum number of iterations and evaluations to 5000 and re-examining the

time taken by glmmTMB, we found FEMA-Long to be up to 7,593 (serial) and 90,265 times faster (parallel) than glmmTMB (**Supplementary Table S3- S4**).

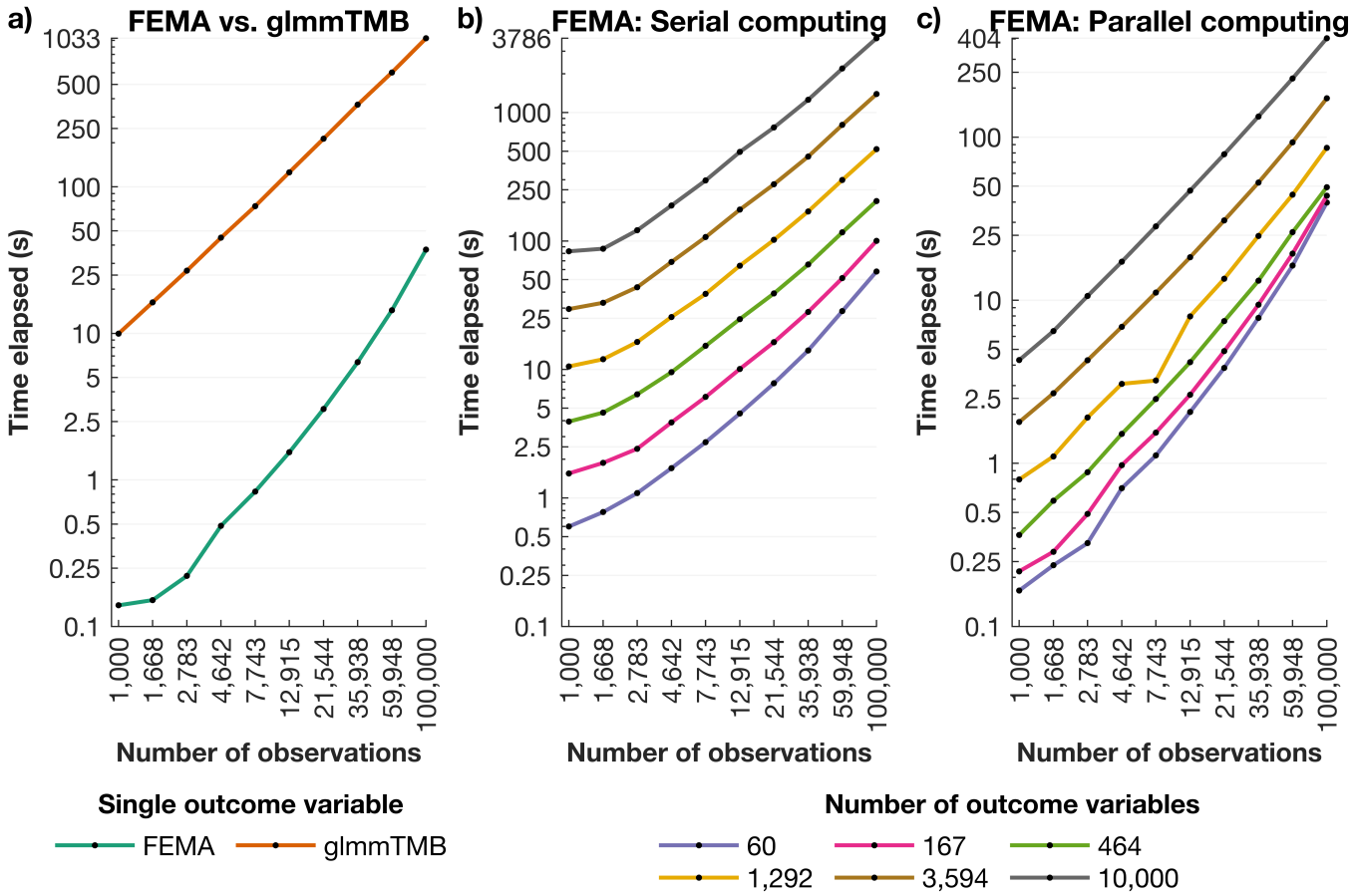

Figure 3: Time taken for fitting linear mixed-effects models with unstructured covariances. In each panel, the line(s) indicate the time taken (in seconds) for fitting linear mixed-effects models with unstructured covariances for family and subject random effects and 50 fixed effects across a range of number of observations (shown on the *x*-axis). Panel **a)** shows the time taken by FEMA (green) and by glmmTMB (orange) for a single outcome variable across a range of observations. Panel **b)** shows the time taken by FEMA with serial computing and panel **c)** shows the time taken by FEMA with parallel computing (32 parallel processes) across a range of increasing number of outcome variables, indicated by the color of the line. Note that the *x*- and *y*-axes are log-spaced in all the panels.

Examining unstructured covariances and discovery of age-dependent genetic effects

To demonstrate the features of FEMA-Long (unstructured covariance, non-linear modeling, scalability to large datasets, performing GWAS, and discovery of time-dependent genetic effects), we performed a GWAS on the length, weight, and BMI of infants during the first year of life using the Norwegian Mother, Father and Child Cohort Study (MoBa)[13,14]. The sample consisted of 68,273 children (46,453 families) with up to six measurements per child (299,442 observations): birth, six weeks, and three, six, eight, and twelve months (see supplementary material for details). We standardized the phenotypes (zero mean, unit standard deviation) and included 51 covariates: intercept, six spline basis functions of age, dummy-coded sex, twenty genetic principal components (PC), and 23 dummy-coded genotyping batches. We created the spline basis functions by placing a knot at the median age for each timepoint: 44, 94, 181, 246, and 369 days, and boundary knots at the extremes (0 and 425 days). The random effects included unstructured covariances for

the family effect, GRM, and the subject effect. Specifically, the stage-1 model (without GWAS) for each phenotype was:

$$\text{Phenotype} \sim 1 + s(\text{age}) + \text{sex} + \text{PC}_{1-20} + \text{Batch} + \text{us}(1|\text{Family}) + \text{us}(1|\text{GRM}) + \text{us}(1|\text{Subject}) \quad (1)$$

We performed GWAS on 6,981,748 SNPs[15], and examined the interaction of each SNP with the six basis functions and the intercept, resulting in seven coefficients per SNP per phenotype. Specifically, after fitting the stage-1 model and residualizing each genotype for the fixed effects in (1), the model for each residual SNP, G_{res}, was:

$$\text{Phenotype}_{\text{residuals,GLS}} \sim G_{\text{res}} + G_{\text{res}} \odot s(\text{age}) \quad (2)$$

where $G_{\text{res}} \odot s(\text{age})$ indicates the element-wise multiplication of G_{res} with every column of the basis function matrix. Then, for each SNP, we performed a Wald test to determine if any of these coefficients had a non-zero effect. Additionally, we also ran a standard longitudinal GWAS, without including $s(\text{age})$ and $G_{\text{res}} \odot s(\text{age})$ terms.

The stage-1 analysis was completed in ~3 minutes; examining the random effects unstructured covariance revealed dynamic patterns over time (**Figure 4**). Specifically, the proportion of within-visit variance attributed to genetic effects (within-visit heritability) for each phenotype showed differences across timepoints, ranging between 0.42–0.91 for length, 0.45–0.84 for weight, and 0.54–0.73 for BMI (diagonal entries, middle column, **Figure 4**). Interestingly, genetic correlations between timepoints (off-diagonal entries, middle column, **Figure 4)** also showed fluctuations over time with different patterns across phenotypes. For example, the correlations between birth and other timepoints for length and weight were fairly stable, while BMI showed a decaying pattern. The heritability was relatively lower for the first three timepoints for length and weight compared to other timepoints, while the heritability for BMI was relatively stable from 3m to 12m period. The family effect explained a small amount of variance for the first three timepoints for length and weight, but not as much for BMI. The variances explained by subject effect (and noise) also changed over time: a stronger effect for the first three timepoints for length but not as much for later timepoints, and a similar trend for weight and BMI but with greater variance explained for later timepoints.

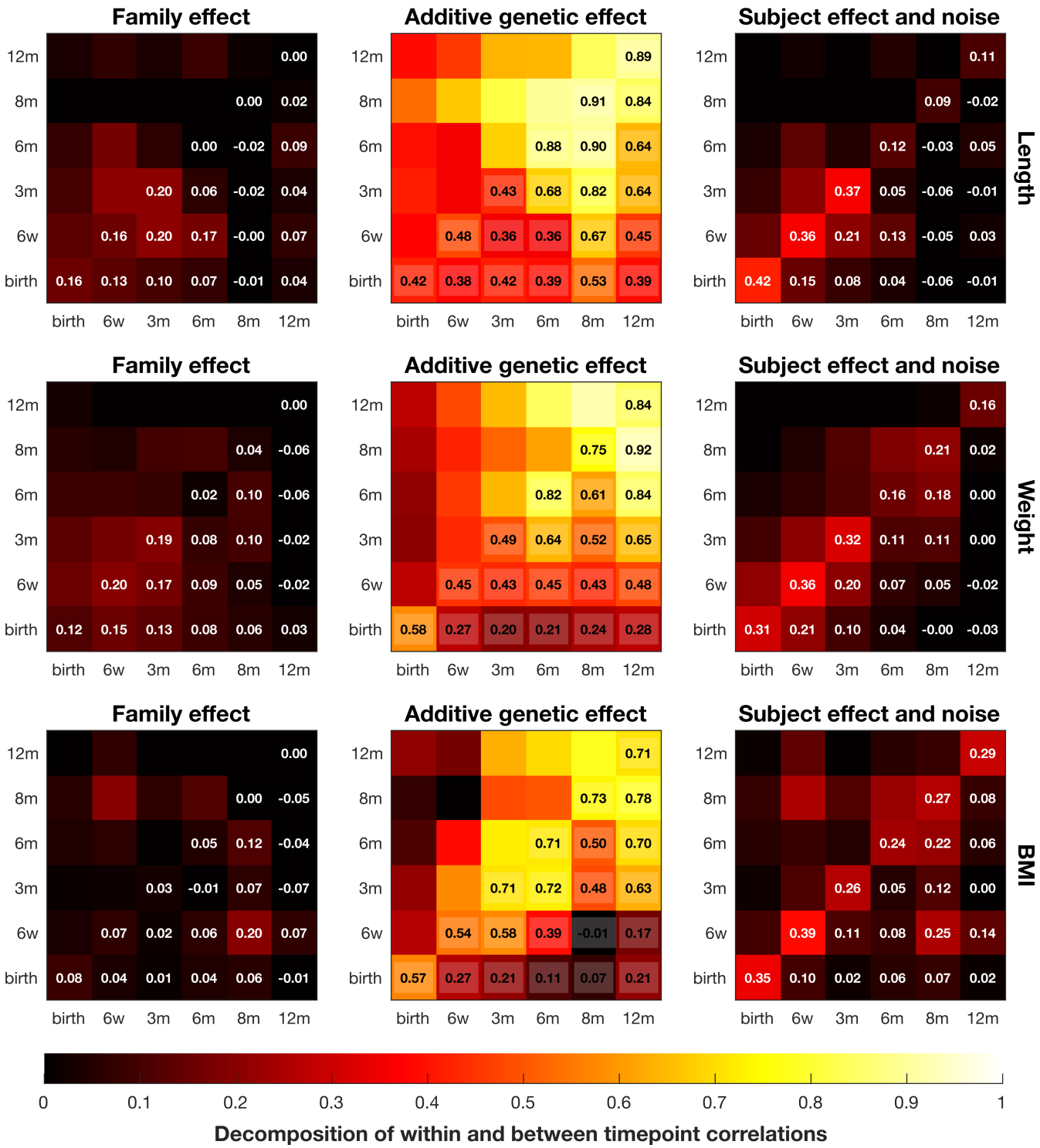

Figure 4: Decomposition of within and between timepoint correlations into family, additive genetic, and subject effects. The panels are the normalized unstructured covariance matrices for length, weight, and BMI in infants in the first year of life; the diagonal values are the variances explained by shared family environment, the additive genetic effect (heritability), and the subject effect (which also includes noise) for each timepoint while the off-diagonals reflect the correlations between time points.

The GWAS analysis, completed in ~16.6 hours, revealed 2,825, 4,383, and 6,664 SNPs showing a main effect or a time-dependent effect on length, weight, and BMI respectively at $\alpha = 5 \times 10^{-8}$; in comparison, longitudinal GWAS (without including splines or interactions; analysis completed in ~3 hours) revealed 2,125, 659, and 2,795 significant SNPs for length, weight, and BMI respectively (**Figure 5a-c**). We examined the fitted spline trajectories for the top 10 SNPs for each trait, one SNP per chromosome (**Figure**

5d). For length, most of the effects were linear; however, the selected SNPs for weight and BMI showed more dynamic patterns with non-linear trajectories. Notably, even the same selected SNPs for weight and BMI showed differing patterns in the direction and timing of the effect changes across the first year of life (e.g., rs2767486 and rs13322435[3]). These trajectories highlight the importance of modeling the time-dependent effects and reveal how the SNP effects change over time.

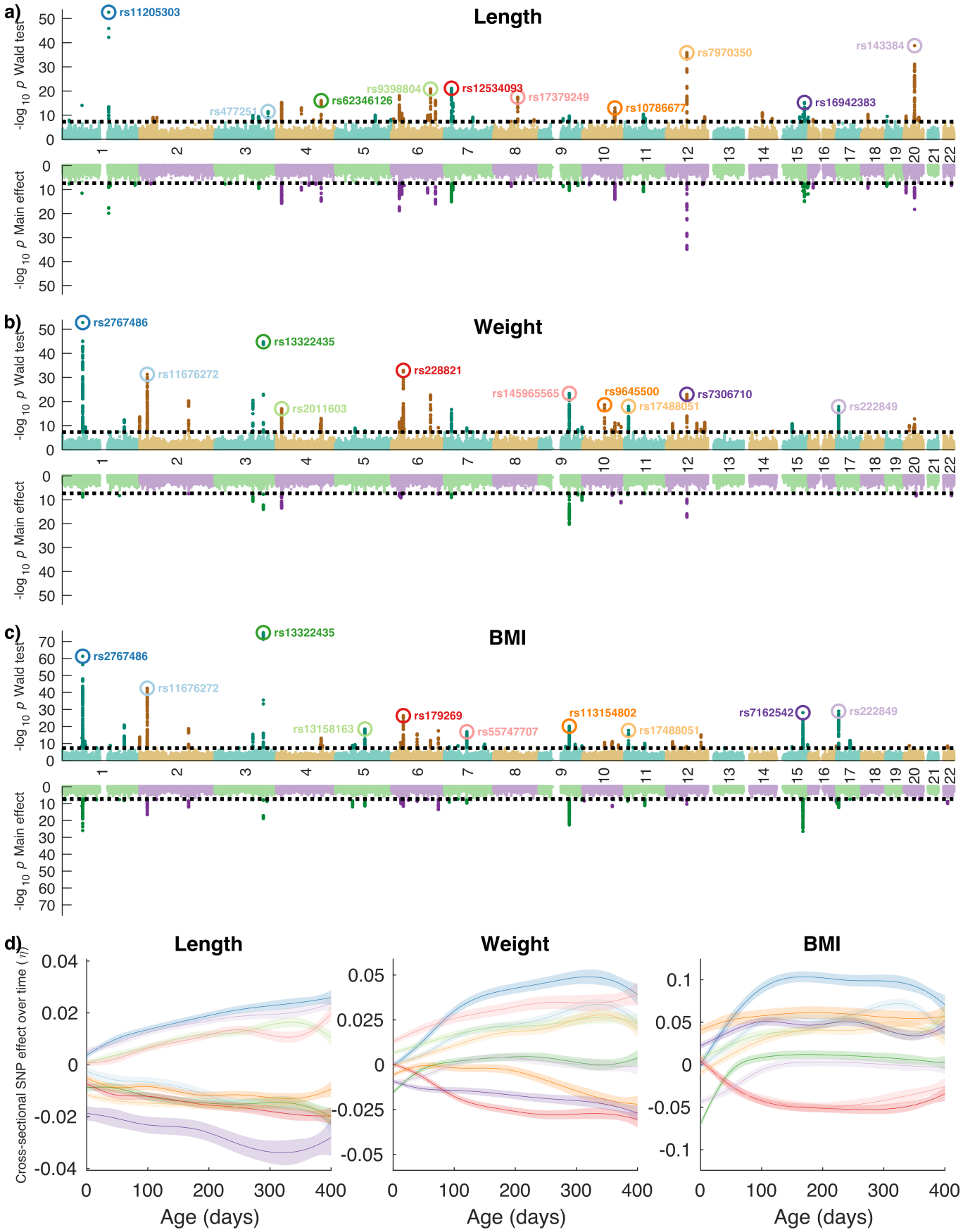

Figure 5: Miami plots comparing the omnibus effect of main and time-dependent effect of SNPs with the longitudinal main effect of SNPs on length, weight, and BMI of infants in the first year of life, and fitted spline trajectories. In panels **a)-c)**, the top part shows the $-\log_{10}(p)$-values for the Wald test while the lower part shows the $-\log_{10}(p)$-values for the longitudinal main effect; in panel **d)** the solid lines show the cross-sectional effect of selected SNPs over time (SNPs selected separately within each phenotype, highlighted in panels a), b), and c)); these are the fitted spline trajectories, showing the

weighted sum of the SNP effect over time; the corresponding standard errors are shown by the band surrounding the solid lines; the colors in panel d) correspond to the colors of the marked SNPs in the corresponding Miami plot.

Discussion

We have adapted and extended our computationally fast linear mixed-effects modeling method FEMA for large-scale high-dimensional longitudinal datasets. The features of FEMA-Long include: i) modeling time-varying random effects using unstructured covariance, ii) incorporating splines for modeling non-linear effects of variables, iii) examining time-dependent effects using spline interactions, and iv) performing GWAS-like analyses by scaling FEMA to estimate marginal effect of millions of variables, including time-dependent effects. Our simulation results show that FEMA-Long can accurately model these parameters, while controlling for false positives, and provides exceptional computational performance, tens to thousands of times faster than a standard LME implementation. We have also demonstrated that FEMA-Long is scalable to scenarios with a large number of observations and outcome variables such as GWAS or whole-brain neuroimaging analyses.

As previously mentioned, random effects like heritability can vary significantly over time[3–8]. Similarly, longitudinal data usually shows some form of time-varying correlation. Our application of FEMA-Long to length, weight, and BMI in infants during the first year of life revealed time-varying random effect covariances including heritability. These patterns also differed between similar phenotypes. Generally, increasing heritability for later timepoints (compared to birth) potentially indicates that the phenotypes at earlier timepoints might be more influenced from other sources like the maternal genotype[16,17]. Importantly, it underscores the need to model these using unstructured covariance as these covariances cannot be well-approximated with simpler covariance patterns like compound symmetry or autoregressive models.

Additionally, when studying dynamic phenomenon such as development, it is essential to model time-dependent effects. Our time-dependent GWAS results revealed several SNPs showing time-dependent genetic effects. In contrast, the discovery of these SNPs was stunted when only examining the longitudinal main effect of SNPs, most notably for weight and BMI. Charting the effect of these SNPs over time revealed several interesting patterns including non-monotonic and curvilinear patterns. These results highlight the two-fold importance of modeling spline interactions for every SNP: first, allowing discovery of SNPs showing non-linear time-dependent effects; and second, ensuring more accurate predictions when using these estimated effect sizes (e.g., polygenic risk scores). Both these aspects cannot be captured by just the main effect of the SNPs and will likely be missed even with linear interaction.

The use of unstructured covariance has received relatively little attention in fields with high dimensional data. For example, within neuroimaging, there are no software packages for modeling unstructured covariances, besides AFNI's *3dLME* and *3dLMEr*[18] which supports compound symmetry and order 1

autoregressive process[9]. In genetics, prior work on longitudinal GWAS include: i) computing a rate of change between timepoints[19–21]; ii) using standard LME packages[22,23]; iii) fitting a growth curve-like model to the phenotype and performing GWAS on the model parameters[24,25]; and iv) novel tools for longitudinal GWAS like fGWAS[26], GALLOP[27], trajGWAS[28], L-BRAT and RGMMAT[29], RVMMAT[30], and SPA_{GRM}[31]. Despite the availability of these tools, they have one or more limitations including inability to: i) handle related individuals, ii) model additional sources of variances, iii) model linear and/or non-linear SNP interactions, iv) estimating effect sizes, and v) scale to a large-sample and/or large-number of phenotypes. Importantly, it is rare to see a longitudinal GWAS study specifying a time-varying covariance pattern. The closest such recent approach[32] used a range of LMEs with linear or cubic smoothing splines, or cubic slopes for age, with and without a continuous AR(1) covariance pattern. However, this work still relied on collapsing repeated observations into a single measurement (followed by a meta-analysis) as well as removing related observations. In addition, the authors highlighted the significant computational burden when scaling to large cohorts.

FEMA-Long addresses all the above-mentioned limitations, providing a unified framework for performing large-scale analyses, modeling unstructured covariances, non-linear effects, and discovery of time-dependent effects, including time-dependent genetic effects. There are, however, some limitations to our work. First, the MoM estimator is not as efficient as maximum likelihood estimator. Second, currently FEMA-Long only supports continuous outcome variables, assuming normally distributed errors. Third, for GWAS, we assume that the individual contribution of SNPs is small and unlikely to impact the covariance estimation of the random effect. Fourth, we estimate the unstructured covariance for each timepoint separately. We could improve this by estimating all parameters jointly; additionally, we could interpolate the visit-wise estimated covariances, resulting in smoothly changing covariances. In addition, the computational time, especially for GWAS, could be improved. Future work will explore these and incorporate additional features like binning the phenotypes, computation of age-specific risk scores, and the use of longitudinal trajectories for prediction, as well as the application of FEMA-Long to longitudinal complex trait phenotypes such as mental health.

To conclude, FEMA-Long is a powerful novel method that will enable researchers to apply large-scale LME models with unstructured covariances as well as discover time-dependent effects. The unified framework of FEMA-Long not only allows inclusion of subjects with a differing number of timepoints but also handles scenarios like sample relatedness and other sources of variance while providing a computationally fast solution for application to genomics, neuroimaging, and other fields. We expect FEMA-Long to facilitate discoveries in the growing number of large longitudinal datasets becoming available.

Methods

Nomenclature

The term *phenotype* refers to outcome variable or the dependent variable; *fixed effects* refer to independent variables, inclusive of covariates of interest and covariates of no interest; and *random effects* refer to various known sources of nesting and similarity between observations in the data. For longitudinal data, the term *subject* or *participant* refers to a unique individual who is followed over time while the term *observation* refers to one measurement for a given subject. *Visit* or *timepoint* corresponds to an ordered set of repeated measurements for the same subject; for example, *visit 1* means the first measurement (first observation) on the subject, with increasing visit numbers indicating subsequent measurements.

Overview

Given a set of phenotype(s), fixed effects, and random effects, the goal is to estimate the fixed effect coefficients while accounting for random effects. In the original FEMA framework[10], we first calculate the ordinary least squares (OLS) residuals of the phenotype; then, we use them to estimate the variance components of the random effects using method of moments (MoM) or a regression estimator with a non-negativity constraint[33]. Then, the fixed effects coefficients are estimated using a generalized least squares (GLS) solution which incorporates a covariance matrix using the estimated variance components of the random effects. For scalability to large number of variables (such as for voxel-wise or vertex-wise data derived from brain imaging data), we use a binning approach – for all outcome variables that have similar variance components for the random effects, the GLS estimation is performed using a common covariance matrix that relies on the average of the variance components of all the outcome variables within that bin.

Here, we present FEMA-Long, an extension of FEMA that is adapted and optimized for large longitudinal datasets. The FEMA-Long framework includes i) time-varying covariances in the random effects using unstructured covariances; ii) modeling non-linear effects using splines; iii) examining time-dependent effects using spline interactions; and iv) performing GWAS and other similar omics analyses, where we test the association of a large number of variables of interest (e.g., genotyping matrix with millions of SNPs) with outcome(s), independently (i.e., marginal rather than joint effect), after controlling for covariates and random effects.

In FEMA-Long, we use unstructured covariance to model time-varying covariance of random effects. The elements of the unstructured covariance matrix are estimated by using the MoM estimation for every pair of visits that are nested within subjects. This results in a matrix of covariance terms for each random effect,

which is then used for GLS estimation of the fixed effects. Importantly, this method does not assume a balanced design of all subjects having data for all the visits.

FEMA-Long also includes the option to model continuous outcome variables as a smooth function (for example, smooth function of age) using splines, represented by a set of basis functions. We also allow interaction terms between fixed effects, including those effects that are modeled via splines. While the estimation of the unstructured covariance matrix uses pre-defined set of timepoints (visits), the spline modeling uses a continuous variable (such as age). This, therefore, brings the framework of generalized additive mixed models to FEMA, allowing flexible modeling of outcome variables as well as performing non-linear interaction analyses.

Finally, while the original FEMA was designed for many outcome variables and a common model (mass-univariate analyses), FEMA-Long includes a variant of FEMA that is intended for GWAS, where we fit the outcome(s) repeatedly on many models, each differing only in their genetic covariate. For this, we leverage the Frisch–Waugh–Lovell theorem[34,35]: first, use stage-1 FEMA to estimate the random effects variance components and fixed effects (omitting genetic variants); then, we GLS residualize genotype vector for the fixed effects using the stage-1 variance components. Finally, we perform a GLS estimation using the GLS residuals of the phenotype and the GLS residuals of the genotype vector. This framework also allows modeling the interaction of genetic variants with spline basis functions, thereby allowing discovery of genetic variants that may show a nonlinear, time-dependent effect on the phenotype.

In the following sections, we describe the details of each of these steps.

General FEMA model set up (stage-1)

Let a dataset have n observations indexed with $i \in 1 \ldots n$. Let X be a full-rank matrix of p fixed effects, with each column of X being a different covariate. X_i are the values of the fixed effects for observation i (X_i is a vector if $p = 1$ and a matrix if $p > 1$). The phenotype is indicated by y, thus y_i indicates the value of the phenotype for observation i. For brevity, we present the framework for a single phenotype, although FEMA fits the same model for each phenotype.

The linear model setup is:

$$y_i = X_i\beta + \varepsilon_i \tag{3}$$

where β is a vector of p weights associated with each fixed effect, and ε_i is the unmodeled measurement error for each observation i. In a mixed modeling framework, we assume that the error term, ε, is normally distributed with zero mean and variance V:

$$\varepsilon \sim \mathcal{N}(0, V) \tag{4}$$

Then, the total residual variance is $\sigma^2 = \frac{\mathrm{tr}(V)}{n}$. This normalized residual variance can be decomposed as a combination of random effects:

$$\frac{V}{\sigma^2} = \sigma_F^2 Z_F Z_F^\top + \sigma_A^2 Z_A Z_A^\top + \sigma_S^2 Z_S Z_S^\top + \sigma_\epsilon^2 I \tag{5}$$

where, generically, Z is a random effects indicator matrix with rows of Z_i indicating membership of observation i; for example, in equation (5), Z_F represents family with a value of 1 indicating observations that belong to the same family, entries of Z_A represent the genetic relationship between a pair of observations, Z_S represents subject effect with a value of 1 indicating repeated measurements for the same subject, and I represents the independent unmodeled variance across each observation; each random effect is associated with a corresponding variance component σ^2. The choice of which random effects should be included depends on the data.

Within FEMA, we first use OLS to get an estimate of the fixed effects coefficients:

$$\hat{\beta} = (X^\top X)^{-1} X^\top y \tag{6}$$

Then, the OLS residuals are:

$$\hat{y}_i^{\mathrm{res,OLS}} = y_i - X_i \hat{\beta} \tag{7}$$

and the total residual variance $\hat{\sigma}^2$ estimated as:

$$\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^{n} \left(\hat{y}_i^{\mathrm{res,OLS}}\right)^2 \tag{8}$$

Following (5), for each pair of observations $\{i, i'\}$, the expected value of the product of corresponding residuals $\hat{y}_i^{\mathrm{res,OLS}}$ and $\hat{y}_{i'}^{\mathrm{res,OLS}}$ is:

$$E\,\frac{\left(\hat{y}_i^{\mathrm{res,OLS}} \times \hat{y}_{i'}^{\mathrm{res,OLS}}\right)}{\hat{\sigma}^2} \approx \sigma_F^2 \left((Z_F Z_F^\top)_{\{i,i'\}}\right) + \sigma_A^2 \left((Z_A Z_A^\top)_{\{i,i'\}}\right) + \sigma_S^2 \left((Z_S Z_S^\top)_{\{i,i'\}}\right) + \sigma_\epsilon^2 \tag{9}$$

where the E operator is the expectation with respect to the random effects and generically $(ZZ^\top)_{\{i,i'\}}$ denotes the $\{i, i'\}$ element of the random effects design matrix. Solving (9) results in the estimated values of random effects variance components, which can be used to construct a covariance matrix $\hat{V}$:

$$\hat{V} = \hat{\sigma}_F^2 Z_F Z_F^\top + \hat{\sigma}_A^2 Z_A Z_A^\top + \hat{\sigma}_S^2 Z_S Z_S^\top + \hat{\sigma}_\varepsilon^2 I \tag{10}$$

Finally, the stage-1 fixed effects can be estimated using the GLS solution:

$$\hat{\beta} = \left(X^\top \hat{V}^{-1} X\right)^{-1} X^\top \hat{V}^{-1} y \tag{11}$$

with variance

$$\mathrm{Var}\left(\hat{\beta}\right) = \left(X^\top \hat{V}^{-1} X\right)^{-1} \tag{12}$$

More details of these steps can be found in[10]. We refer to this overall procedure as stage-1, which includes estimation of the random effects parameters as well as GLS estimation of the fixed effects coefficients.

Estimating unstructured covariance

The original framework in FEMA assumed compound symmetry (i.e., constant covariance over time), an assumption which may not hold for longitudinal studies. In FEMA-Long, we allow modeling arbitrary unstructured covariance, where the parameters of the unstructured covariance matrix are estimated using the available data. Specifically, using the OLS residuals from equation (7), we solve equation (13) for every pair of visits (v, v'):

$$E\,\frac{\left(\hat{y}_i^{\mathrm{res,OLS}} \times \hat{y}_{i'}^{\mathrm{res,OLS}}\right)_{vv'}}{\hat{\sigma}^2} \approx \Omega_{F,vv'}\left(\left(Z_F Z_F^\top\right)_{\{i,i'\}}\right)_{vv'} + \Omega_{A,vv'}\left(\left(Z_A Z_A^\top\right)_{\{i,i'\}}\right)_{vv'} + \Omega_{S,vv'}\left(\left(Z_S Z_S^\top\right)_{\{i,i'\}}\right)_{vv'} \tag{13}$$

When solving (13), if $v = v'$, we use non-negativity constrained solution to estimate the variance component for that visit number v; if $v \neq v'$, we drop the non-negativity constraint and estimate the covariance of the random effects between pairs of visits v and v'. Then, the estimated variance-covariance components can be put together to get $v \times v$ variance-covariance matrix for each random effect. Note that since these terms are being estimated within each pair of visits, the subject and error variances ($\hat{\sigma}^2_{S,vv}$ and $\hat{\sigma}^2_{\epsilon,vv}$) cannot be uniquely identified, and thus are estimated as one. This formulation does not require a balanced dataset – i.e., subjects with differing number of visits are allowed and the covariances are estimated according to the available data, with no observations being removed.

After estimating the elements of the unstructured covariance matrix, we construct the covariance matrix $\hat{V}$ following equation (10)), with appropriate visit-wise reordering (such that the correct variance or covariance term for each random effect is used). Then, we use this covariance matrix $\hat{V}$ to perform GLS estimation for the fixed effects (equation (11)).

Modeling spline basis functions and omnibus test

FEMA allows modeling the outcome as a set of smooth basis functions $f_s(t)$ (typically splines) of continuous variable t:

$$y_i = X_i\beta + \sum_{s=1}^{S} \beta_s f_s(t) + \varepsilon_i \tag{14}$$

This leads to representing the variable t by S derived columns in the design matrix X. Then, the set of coefficients, β_s, define the non-linear function $\mathcal{F}(t) = \sum_{s=1}^{S} \beta_s f_s(t)$, which represents the trajectory of the outcome over, for example, time. These basis functions can also be used to model interactions with another fixed effect.

In FEMA, the default option is to create natural cubic splines with unit height at knots, similar to the *nsk* option in the *splines2* package[36,37] in *R*. Users can specify the knot placement based on domain knowledge, experiment design, or choose to place the knots at the quartiles of the variable. We also support creating other spline basis functions, namely natural cubic splines or B-splines which internally rely on a call to *ns* or *bs* functions in the *splines* package[12] in *R*.

By default, we modify the created basis functions such that the linear (or main) effect of t can be retained in the model. First, we create a linearly spaced vector of values t_{lin} over the range of values in t. Then, we create basis functions using these values. At this point, the user has an option of regressing out powers of t_{lin} – for example, regressing out the zeroth and first powers of t_{lin} would result in demeaned basis functions that are orthogonal to the linear effect of t_{lin}. Next, we perform a singular value decomposition on the demeaned basis functions. We then rescale the orthonormal basis functions such that each basis function has a bounded value between $[-1, 1]$. Finally, we linearly interpolate the modified basis functions to get corresponding basis function values for t. These modified orthonormal basis functions are added as columns of X, in addition to retaining t as an independent variable.

Once the coefficients have been estimated, we can perform an omnibus Wald test to test the null hypothesis that the linear combination of the estimated coefficients is zero. Generically, given a vector or matrix of weights L with rank r equal to number of linearly independent rows, for a vector of estimated coefficients $\hat{\beta}$ the multivariate Wald statistics W is[38]:

$$W = \left(L\hat{\beta}\right)^{\top}\left\{L\widehat{\text{Cov}}(\hat{\beta})L^{\top}\right\}^{-1}\left(L\hat{\beta}\right) \tag{15}$$

where $\widehat{\text{Cov}}(\beta)$ is the estimated variance-covariance matrix of the fixed effects coefficients. The Wald statistic W follows a χ^2 distribution with r degrees of freedom. Therefore, for an omnibus test of the null

hypothesis that the linear combination of estimated coefficients of the basis functions is zero, $L = I(S)$, i.e., an identity matrix of size S (the number of basis functions). Alternatively, it is possible to use an F distribution to calculate the p value with two degrees of freedom: the numerator degree of freedom is the rank of L and the denominator degree of freedom is $n - p - S$ (see[39] for a discussion on the similarity and differences in these approaches).

Scaling FEMA for GWAS-like analyses

Within GWAS, the goal is to estimate the effect of genetic variants on the phenotype(s), considering one variant at a time. Let G represent the time-invariant genotyping matrix, in which entry $g_i \in \{0, 1, 2\}$ denotes the best-guess genotype (imputed genotypes which have a posterior probability above a threshold, followed by conversion to discrete count indicating the most probable genotype[40]) for the single nucleotide polymorphism (SNP) for observation i; 0 represents homozygous for the reference allele, 1 represents heterozygous, and 2 represents homozygous for the other allele. Let γ_g be the weight associated with this SNP g. Since we are only interested in the marginal effect of each SNP g in G, instead of fitting the full model (with covariates) for every SNP, we employ a two-stage procedure.

The Frisch-Waugh-Lovell theorem[34,35] states that the estimate of coefficients from a full regression model are equivalent to regression done in parts (or partial regressions) if both the X and y variables are residualized[41]. This implies that we can estimate γ_g by first fitting a reduced model without genetic effect (stage-1), then residualizing each genetic variant g for the fixed effects, and finally fitting a second model using the residuals from stage 1 as the dependent variable and the residualized genotype as the independent variable (see[42] for a proof of the Frisch-Waugh-Lovell theorem in the GLS case).

Concretely, we first fit a reduced model excluding the genotyping vector. We estimate the variance components (compound symmetry or unstructured covariance) and the GLS estimates for the fixed effects using the stage-1 model as described previously. Then, the GLS residuals can be calculated as:

$$\hat{y}_i^{\text{res,GLS}} = y_i - X_i\hat{\beta} \tag{16}$$

Next, we residualize g for the fixed effects. The coefficients of X on g, ζ, can be calculated as:

$$\hat{\zeta} = \left(X^\top \hat{V}^{-1} X\right)^{-1} X^\top \hat{V}^{-1} g \tag{17}$$

resulting in the GLS residuals:

$$\hat{g}_i^{\text{res,GLS}} = g_i - X_i\hat{\zeta} \tag{18}$$

It is important to note that this residualized genotype vector $\hat{g}^{\text{res,GLS}}$ is phenotype specific. This follows from $\hat{V}$ being composed from phenotype specific variance components (equation (10)). Now, using the stage-1 residuals $\hat{y}_i^{\text{res,GLS}}$ and the residualized genotype $\hat{g}_i^{\text{res,GLS}}$, we can estimate $\hat{\gamma}_g$ as:

$$\hat{\gamma}_g = \left(\hat{g}^{\text{res,GLS}\top}\hat{V}^{-1}\hat{g}^{\text{res,GLS}}\right)^{-1}\hat{g}^{\text{res,GLS}\top}\hat{V}^{-1}\hat{y}^{\text{res,GLS}} \tag{19}$$

Then, for each genetic variant (and for each phenotype), the GLS residuals can be calculated as:

$$\tilde{y}_{i,g}^{\text{res,GLS}} = \hat{y}_i^{\text{res,GLS}} - \hat{g}_i^{\text{res,GLS}}\hat{\gamma}_g \tag{20}$$

We use the notation $\tilde{y}_{i,g}^{\text{res,GLS}}$ to distinguish these residuals (which are specific to each genetic variant) from $\hat{y}_i^{\text{res,GLS}}$ which are the stage-1 GLS residuals obtained from FEMA after accounting for the fixed effects (without the genetic variant). In essence, the residuals $\hat{y}^{\text{res,GLS}}$ from (16) are the residualized phenotype for (19) which are estimated using the residualized genotype $\hat{g}^{\text{res,GLS}}$.

The mean squared error is given by:

$$\tilde{\sigma}_g^2 = \frac{1}{(n-p-1)}\sum_{i=1}^{n}\left(\tilde{y}_{i,g}^{\text{res,GLS}}\right)^2 \tag{21}$$

where the denominator is normalized by $(n-p-1)$ to account for the genotyping vector g as an additional independent variable[35,43]. The variance for each $\hat{\gamma}_g$ can be estimated as:

$$\text{Var}(\hat{\gamma}_g) = \left(\hat{g}^{\text{res,GLS}\top}\hat{V}^{-1}\hat{g}^{\text{res,GLS}}\right)^{-1} \tag{22}$$

Since the $\hat{V}$ term has been normalized by $\hat{\sigma}^2$, the standard error is scaled by $\tilde{\sigma}_g^2$.

The above framework can be applied to other GWAS-like analyses where one is interested in examining the effect of many predictors, taken one at a time (for example, given protein assays, the effect of each protein on a phenotype).

Modeling time-dependent genetic effects using spline interactions

This GWAS framework can also include interaction terms – linear interaction or smooth interaction. In case of linear interaction, if the main effect of the interacting variable has already been included as an X variable, g is then a matrix with the first column containing the main effect of the genetic variant and the second column containing the elementwise (Hadamard) product of the genetic variant and the interacting variable. We residualize both columns for X (equation (17)), followed by estimation using equation (19), resulting

in two coefficients: the coefficient for the main effect and the coefficient for the interaction term. The denominator of the mean squared error (equation (21)) is adjusted for two parameters, i.e., $(n - p - 2)$.

For modeling spline interactions, g is a matrix having the elementwise product of the genetic variant and the smooth function $f_s(t)$. The estimation follows the same procedure, with the estimation of as many coefficients as the columns of g. If the basis functions have been appropriately modified and a constant term added (see the section on Modeling spline basis functions), then both the main effect of the genetic variant as well as the interaction of the genetic variant with smooth functions can be estimated. Generically, if there are S basis functions (with or without the constant term), the denominator in equation (21) is penalized by $(n - p - S)$.

Concretely, the GWAS interaction can be framed as modeling the effects of g on the outcome by altering regression coefficients β_s in (14):

$$y_i = X_i\beta + \sum_{s=0}^{S}(\beta_s + \gamma_s g)f_s(t) + \varepsilon_i \tag{23}$$

where γ_s coefficients are equivalent to the interaction term between variable g and the basis functions $f_s(t)$. Note that $s = 0$ corresponds to the main effect of the SNP. This formulation is equivalent to:

$$y_i = X_i\beta + \sum_{s=1}^{S}\beta_s f_s(t) + \gamma_g g + \sum_{s=1}^{S}\gamma_s g f_s(t) + \varepsilon_i \tag{24}$$

For example, if the time t was modeled as smooth functions $f_s(t)$, then after estimating the regression coefficients, FEMA can be used to calculate the following quantities:

The cross-sectional effect of the genotype g on the outcome, at a specific time t_0:

$$\eta_{t_0} = \frac{dy}{dg} = \sum_{s=0}^{S}\hat{\gamma}_s f_s(t_0) \tag{25}$$

and the instantaneous effect of genotype g on the rate of change of the outcome at time t_0:

$$\eta'_{t_0} = \frac{d^2y}{dgdt} = \sum_{s=1}^{S}\hat{\gamma}_s f'_s(t_0) \tag{26}$$

Both these quantities can be plotted as a function of time, t, to assess the effect of g on the outcome over time.

Simulations

To validate the FEMA approach for fitting LMEs with unstructured covariances, we performed a series of simulations. All analyses were performed using MATLAB R2023a[44].

Simulation 1: parameter recovery

In the first simulation, we created phenotypes with known variance-covariance matrices and known effect sizes. Then, using FEMA, we estimated the parameters and compared them with the ground truth. We created a grid of known sources of variances: fixed effects $\mathrm{Var}_{\mathrm{FFX}} \in [0.1,0.9]$, family effect (i.e., variance attributed to common environment) $\mathrm{Var}_{\mathrm{Fam}} \in [0.1,0.9]$, subject effect (i.e., variance attributed to repeated measurements) $\mathrm{Var}_{\mathrm{Sub}} \in [0.1,0.9]$, and noise $\mathrm{Var}_{\mathrm{Noise}} \in [0.1,0.9]$. The total phenotypic variance was the sum of these four variances with the total variance being 1; i.e., $\mathrm{Var}_y = \mathrm{Var}_{\mathrm{FFX}} + \mathrm{Var}_{\mathrm{Fam}} + \mathrm{Var}_{\mathrm{Sub}} + \mathrm{Var}_{\mathrm{Noise}} = 1$. This resulted in 84 combinations of variances.

For each of these combinations, we simulated data for 5,000 unique individuals having up to five repeated measurements and up to five members within a family, with a total of 20,000 observations. We simulated the age at each measurement using the mean and standard deviation of age as $\{0 \pm 0;\ 60 \pm 10;\ 180 \pm 12;\ 240 \pm 14;\ 365 \pm 15\}$ days – this scenario represents repeated measures data in infants with measurements conducted at birth, two months, six months, eight months, and one year.

For each of the 84 combinations of variances, we created an arbitrary 5×5 covariance matrix for family effect ($\mathrm{V}_{\mathrm{Fam}}$) and subject effect ($\mathrm{V}_{\mathrm{Sub}}$). We simulated 15 fixed effects – intercept, spline basis functions of the simulated age, and other random X variables drawn from a normal distribution. The fixed effects parameters were drawn from a uniform distribution $\beta^{\mathrm{sim}} \sim \mathcal{U}[-0.2, 0.2]$ such that $\mathrm{y}_{\mathrm{FFX}} = X\beta^{\mathrm{sim}}$, having mean $\overline{y_{\mathrm{FFX}}}$ and standard deviation σ_{FFX}; we then mean centered and scaled y_{FFX} to have $\mathrm{Var}_{\mathrm{FFX}}$ variance. Then, using V_{Fam} and V_{Sub} covariance matrices, we drew samples from a multivariate normal distribution with zero means and scaled them to have $\sqrt{\frac{\mathrm{Var}_{\mathrm{Fam}}}{\mathrm{Var}_{\mathrm{RFX}}}}$ and $\sqrt{\frac{\mathrm{Var}_{\mathrm{Sub}}}{\mathrm{Var}_{\mathrm{RFX}}}}$ variances each, where $\mathrm{Var}_{\mathrm{RFX}} = \mathrm{Var}_{\mathrm{Fam}} + \mathrm{Var}_{\mathrm{Sub}}$. In other words, $y_{\mathrm{Fam}} = \sqrt{\frac{\mathrm{Var}_{\mathrm{Fam}}}{\mathrm{Var}_{\mathrm{RFX}}}} \times \mathrm{MVN}(0, V_{\mathrm{Fam}})$ and $y_{\mathrm{Sub}} = \sqrt{\frac{\mathrm{Var}_{\mathrm{Sub}}}{\mathrm{Var}_{\mathrm{RFX}}}} \times \mathrm{MVN}(0, V_{\mathrm{Sub}})$. Therefore, $y_{\mathrm{RFX}} = \sqrt{\mathrm{Var}_{\mathrm{RFX}}} \times (y_{\mathrm{Fam}} + y_{\mathrm{Sub}})$. Finally, we simulated noise drawn from a uniform distribution and scaled it to have zero mean and $\mathrm{Var}_{\mathrm{Noise}}$ variance. Thus, the simulated phenotype was $y = y_{\mathrm{FFX}} + y_{\mathrm{RFX}} + y_{\mathrm{noise}}$.

For this simulated phenotype, we fitted the model using FEMA, specifying the covariance type to be unstructured and the random effects to be family and subject effects. As an additional comparator, we fit the same model using the *glmmTMB* toolbox version 1.1.10[11] in R version 4.2.1[12]. The model in glmmTMB was specified as: $y \sim 1 + X + \mathrm{us}(\mathrm{timepoint} + 0|\mathrm{FID}) + \mathrm{us}(\mathrm{timepoint} + 0|\mathrm{IID})$, dispformula = ~0,

where 1 indicates the intercept, X indicates the remaining covariates (i.e., excluding the intercept), us specifics an unstructured covariance matrix for family (FID) and subject (IID) terms, the timepoint denotes a factor variable indicating which visit the observation belonged to, and the dispformula component disables the addition of an additional error term as that would be an overparameterized model[45]. After preliminary testing, we noticed convergence issues with glmmTMB; therefore, we changed the default settings to have higher number of iterations, specifically setting iter.max and eval.max to 5,000.

We compared the beta coefficients for the fixed effects and the estimated variance-covariance matrices for the family and subject effects from both FEMA and glmmTMB with the ground truth and with each other. For the beta coefficients, $\beta_{\text{intercept}}^{\text{true}} = \beta_{\text{intercept}}^{\text{sim}} - \overline{y_{\text{FFX}}}/\sigma_{\text{FFX}}$, and for the remaining X variables, $\beta_X^{\text{true}} = \beta_X^{\text{sim}}/\sigma_{\text{FFX}}$; for the family effect variance-covariance matrix, $V_{\text{Fam}}^{\text{true}} = V_{\text{Fam}} \times \text{Var}_{\text{Fam}}$; and for the subject effect variance-covariance matrix, $V_{\text{Sub}}^{\text{true}} = V_{\text{Sub}} \times \text{Var}_{\text{Sub}} + I \times \text{Var}_{\text{noise}}$, where I was an identity matrix of size five (corresponding to number of visits). Then, we calculated the root mean squared error (RMSE) for the fixed effects and the family and subject variance-covariance matrices to assess how far the estimated values were from the true simulated parameters, and additionally to quantify the difference between the estimated parameters from FEMA and glmmTMB. For each of the 84 combinations of variances explained by fixed and random effects, we repeated this entire process 50 times.

Simulation 2: false positive rate

To evaluate the false positive rate, we repeated the strategy from simulation one. The fixed effects included the intercept, age, and other random 98 X variables drawn from a random normal distribution, resulting in 100 X variables. However, the fixed effects had zero contribution to the simulated phenotype. The total phenotypic variance was 1, i.e., $\text{Var}_y = \text{Var}_{\text{Fam}} + \text{Var}_{\text{Sub}} + \text{Var}_{\text{Noise}} = 1$, resulting in 36 different combinations. For each simulation setting, we created 10 y variables that have different amount of contribution from family, subject, and noise terms and repeated these simulations 1,000 times. Using these simulated data, we fitted the models using the unstructured covariance in FEMA as well as repeated the model fitting using compound symmetry covariance type. Then, we examined the distribution of the p-values from both the models.

Simulation 3: computational performance

To assess the computational performance of FEMA with unstructured covariance, we simulated data with varying number of observations and varying number of outcome variables. We fixed $\text{Var}_{\text{Fam}} = 0.20$, $\text{Var}_{\text{Sub}} = 0.30$, $\text{Var}_{\text{FFX}} = 0.35$, and $\text{Var}_{\text{noise}} = 0.15$. We set the number of fixed effects to 50. We created 10 log-spaced values for the number of observations (rounded off), varying between 10^3 and 10^5. Similarly, we created 10 log-spaced values for the number of outcome variables (rounded off), varying between 10^0

and 10^4. For each of these combinations of number of observations and number of outcome variables, we estimated the time taken by FEMA. We repeated the time measurement five time and then took the median to get a robust estimate of the time taken to run the analysis. We also performed the time measurement using parallel computing with 24 parallel jobs. In addition, we calculated the time taken by glmmTMB to run each of these models across the 10 values of number of observations. Since glmmTMB is not designed for multiple outcome variables, we only timed it for one outcome variable and then scaled the timing for larger number of outcome variables. We ran the timing measurements on a compute cluster with AMD EPYC 7702 processor with 64 cores, with the job specifications of 64 CPUs per task and 7GB memory per CPU.

Application: GWAS on anthropometric features in infants

Study sample

To demonstrate the use of FEMA, we performed a GWAS on the length, weight, and BMI of infants in the first year of life using the large-scale Norwegian Mother, Father, and Child Cohort Study (MoBa)[13,14]. The MoBa Study is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Participants were recruited from all over Norway from 1999-2008. The women consented to participation in 41% of the pregnancies. The cohort includes approximately 114,500 children, 95,200 mothers and 75,200 fathers. The current study is based on version 12 of the quality-assured data files released for research in 2019. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics (2016/1226/REK Sør-Øst C).

Genetic data

We used the quality controlled genetic data from the MoBaPsychGen pipeline v1[15]. The release consists of data from 76,577 children, 77,634 mothers, and 53,358 fathers; we restricted ourselves to the data on children who had not withdrawn their consent from MoBa as of November 1, 2024.

Phenotypes

The phenotypes of interest were the length, weight, and BMI of infants at six timepoints during the first year of life: birth, six weeks, three months, six months, eight months, and twelve months. The measurements were derived from a combination of Medical Birth Registry of Norway (MBRN) and questionnaires filled by the mothers six and eighteen months after birth. Within the questionnaire filled by mothers when their child was six months old, the questions included a record of the length and weight of

the child at birth, six weeks, three months, and six months. For the latter three, the mother was asked to refer to the child's health card. Within the questionnaire filled by mothers when their child was eighteen months old, the questions included the length and weight of the child at eight months, around one year, and between fifteen to eighteen months. We included measurements up to year one for our analysis. We performed a series of quality checks prior to performing GWAS. These steps are described in detail in the Supplementary Materials. The final sample size for GWAS was 68,273 infants with 299,442 observations having complete data on length, weight, and BMI (i.e., each of the 68,273 subjects had all three measurements); see **Supplementary Figure S8** for distribution of age at each timepoint, and **Supplementary Figures S9-S11** for plots showing the trajectories of length, weight, and BMI.

Genome-wide association study

For the 68,273 infants, with up to six repeated measurements totaling 299,442 observations on length, weight, and BMI during the first year of life, we performed a longitudinal GWAS using FEMA on the autosomes. We standardized the phenotypes to have zero mean and unit standard deviation. The covariates included the intercept, six spline basis functions (after the transformations described in the Modeling spline basis functions section), dummy-coded sex, twenty genetic principal components, and 23 dummy-coded genotyping batch variables, resulting in a total of 51 covariates. We created spline basis functions by placing a knot at the median age for each visit (**Supplementary Figure S12** and **Supplementary Figure S13**). The random effects included the effect of family or shared environment, the genetic relationship matrix (GRM), and the subject effect (or repeated measurements). We defined families using the mother ID of the participants; in addition, for a subset of participants where the mother ID was different, but the father ID was the same, we assigned those individuals to the same family; this resulted in 46,453 unique families.

To calculate GRM, we used a subset of 500,243 SNPs which were directly genotyped in at least one imputation batch or had an imputation INFO score > 0.99785 in all imputation batches; we mean imputed any missing genotyping data and standardized the SNPs (zero mean, unit standard deviation); then, the GRM was calculated as the correlation coefficient between the standardized SNPs. While performing the GWAS, we mean imputed any missing genotyping data and examined both the main effect of the SNPs as well as the interaction of each SNP with the six basis functions, resulting in the estimation of seven coefficients for each phenotype. For each phenotype, we performed a Wald test to examine if any of these seven coefficients had a non-zero effect. To accelerate computing, we split the genetic data into chunks of 5,000 SNPs, parallelly processing twenty chunks. In addition, we also performed a longitudinal GWAS on the same phenotypes (i.e., without including spline interactions).

Code availability

FEMA-Long is available on GitHub at: https://github.com/cmig-research-group/cmig_tools.

Acknowledgements

The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities.

PP and **NP** acknowledge funding support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 801133; **PP**, **NP**, and **DM** acknowledge funding support from the Research Council of Norway grant number 324252. **EF** is supported by the Research Council of Norway (grant #324499). **MV** is supported by the Research Council of Norway (grant #301178), the European Research Council (grant #101171420) and the University of Bergen. **DMS** is supported by the National Institutes of Health Ruth L. Kirschstein Individual Predoctoral NRSA award (5F30DA057078-02). **IES** is supported by the South-Eastern Norway Regional Health Authority (#2020060). **VB** is supported by the EU Horizon 2020 Research and Innovation Programme (RealMent; #964874). **NRB** was supported by the Research Council of Norway (RCN) (#271555/F21). **CM** acknowledges funding from the National Institutes of Mental Health (R00MH132886) and the Brain Behavior Research Foundation (grant number 31876). **JK** was supported by a Marie Skłodowska-Curie Postdoctoral Fellowship under the European Union’s Horizon Europe research and innovation programme (Grant Agreement No.101150746). **SJ** is supported by Helse Vest's Open Research Grant (grants #912250 and F-12144), the Novo Nordisk Foundation (grant NNF20OC0063872) and the Research Council of Norway (grant #315599). **PRN** was supported by grants from the European Research Council (AdG #293574), Stiftelsen Trond Mohn Foundation (Mohn Center of Diabetes Precision Medicine), the University of Bergen, Haukeland University Hospital, the Research Council of Norway (FRIPRO grant #240413), and the Novo Nordisk Foundation (grant #54741). **OF** acknowledges funding support from the

South-Eastern Norway Regional Health Authority (#2022073) and the Research Council of Norway (#324499). **AAS** is supported by the Research Council of Norway (grant #326813). **TEN** acknowledges funding from the National Institutes of Health grant 5U24DA041123. **AMD** acknowledges funding from the National Institutes of Health grants U24DA041123, R01AG076838, U24DA055330, and OT2HL161847.

Parts of this work used the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT, tsd-drift@usit.uio.no), using resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway (NS9703S).

We acknowledge helpful suggestions from Drs. Gleda Kutrolli and Dominic Holland. We thank Dr. Elizabeth Corfield for leading the MoBaPsychGen quality control and imputation pipeline.

Some of the color schemes used for various figures were taken from ColorBrewer (https://colorbrewer2.org/). Some of the figures use the tight_subplot function (Pekka Kumpulainen (2023). tight_subplot(Nh, Nw, gap, marg_h, marg_w) (https://www.mathworks.com/matlabcentral/fileexchange/27991-tight_subplot-nh-nw-gap-marg_h-marg_w), MATLAB Central File Exchange. Retrieved April 5, 2023).

Conflict of interest

Dr. Anders M. Dale is Founding Director, holds equity in CorTechs Labs, Inc. (DBA Cortechs.ai), and serves on its Board of Directors and Scientific Advisory Board. Dr. Dale is the President of J. Craig Venter Institute (JCVI) and is a member of the Board of Trustees of JCVI. He is an unpaid consultant for Oslo University Hospital. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies. Dr. Andreassen has received speaker fees from Lundbeck, Janssen, Otsuka, Lilly, and Sunovion and is a consultant to Cortechs.ai. and Precision Health. Dr. Frei is a consultant to Precision Health. The other authors declare no competing interests.

Author contribution

PP: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Visualization; Writing – original draft; Writing – review & editing

NP: Conceptualization; Methodology; Visualization; Writing – review & editing

DP: Methodology; Software; Writing – review & editing

EF: Data curation; Investigation; Software; Writing – review & editing

MV: Data curation; Writing – review & editing

DMS: Software; Writing – review & editing

AR: Software; Writing – review & editing

PJ: Data curation; Visualization; Writing – review & editing

IES: Resources, Writing – review & editing

VB: Data curation; Writing – review & editing

NRB: Data curation; Writing – review & editing

CCF: Software; Writing – review & editing

CM: Writing – review & editing

JK: Software; Writing – review & editing

RL: Writing – review & editing

DJH: Writing – review & editing

DM: Writing – review & editing

SJ: Writing – review & editing

PRN: Writing – review & editing

TLJ: Resources; Writing – review & editing

WKT: Methodology; Writing – review & editing

OF: Conceptualization; Data curation; Methodology; Software; Supervision; Writing – review & editing

AAS: Conceptualization; Methodology; Software; Supervision; Writing – review & editing

TEN: Conceptualization; Methodology; Software; Supervision; Visualization; Writing – review & editing

OAA: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing

AMD: Conceptualization; Funding acquisition; Methodology; Project administration; Software; Resources; Supervision; Visualization; Writing – review & editing

References

1. Denny, J. C. & Collins, F. S. Precision medicine in 2030—seven ways to transform healthcare. *Cell* **184**, 1415–1419 (2021).

2. Atlas of Longitudinal Datasets. https://atlaslongitudinaldatasets.ac.uk/ (2024).

3. Helgeland, Ø. *et al.* Characterization of the genetic architecture of infant and early childhood body mass index. *Nat Metab* **4**, 344–358 (2022).

4. Jr, T. J. B. The Wilson Effect: The Increase in Heritability of IQ With Age. *Twin Research and Human Genetics* **16**, 923–930 (2013).

5. Brown, W. M. *et al.* Age-stratified heritability estimation in the Framingham Heart Study families. *BMC Genetics* **4**, S32 (2003).

6. Akingbuwa, W. A., Hammerschlag, A. R., Bartels, M. & Middeldorp, C. M. Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry* **61**, 227–242 (2022).

7. Jelenkovic, A. *et al.* Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Sci Rep* **6**, 28496 (2016).

8. Silventoinen, K. *et al.* Genetic regulation of growth from birth to 18 years of age: The Swedish young male twins study. *American Journal of Human Biology* **20**, 292–298 (2008).

9. van der Horn, H. J. *et al.* A cautionary tale on the effects of different covariance structures in linear mixed effects modeling of fMRI data. *Human Brain Mapping* **45**, e26699 (2024).

10. Parekh, P. *et al.* FEMA: Fast and efficient mixed-effects algorithm for large sample whole-brain imaging data. *Human Brain Mapping* **45**, e26579 (2024).

11. Brooks, M. E. *et al.* glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* **9**, 378–400 (2017).

12. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2024).

13. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology* **45**, 382–388 (2016).

14. Magnus, P. *et al.* Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology* **35**, 1146–1150 (2006).

15. Corfield, E. C. *et al.* The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. 2022.06.23.496289 Preprint at https://doi.org/10.1101/2022.06.23.496289 (2024).

16. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet* **51**, 804–814 (2019).

17. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Human Molecular Genetics* **27**, 742–756 (2018).

18. Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S. & Cox, R. W. Linear mixed-effects modeling approach to FMRI group analysis. *NeuroImage* **73**, 176–190 (2013).

19. Brouwer, R. M. *et al.* Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat Neurosci* **25**, 421–432 (2022).

20. Gorski, M. *et al.* Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies. *Kidney International* **102**, 624–639 (2022).

21. Loomis, S. J. *et al.* Genome-wide study of longitudinal brain imaging measures of multiple sclerosis progression across six clinical trials. *Sci Rep* **13**, 14313 (2023).

22. Jin, H. *et al.* Genome-Wide Association Study on Longitudinal Change in Fasting Plasma Glucose in Korean Population. *Diabetes Metab J* **47**, 255–266 (2023).

23. Wendel, B. *et al.* A genome-wide association study of the longitudinal course of executive functions. *Transl Psychiatry* **11**, 1–8 (2021).

24. Bradfield, J. P. *et al.* Trans-ancestral genome-wide association study of longitudinal pubertal height growth and shared heritability with adult health outcomes. *Genome Biology* **25**, 22 (2024).

25. Couto Alves, A. *et al.* GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Science Advances* **5**, eaaw3095 (2019).

26. Das, K. *et al.* A dynamic model for genome-wide association studies. *Hum Genet* **129**, 629–639 (2011).

27. Sikorska, K., Lesaffre, E., Groenen, P. J. F., Rivadeneira, F. & Eilers, P. H. C. Genome-wide Analysis of Large-scale Longitudinal Outcomes using Penalization —GALLOP algorithm. *Sci Rep* **8**, 6815 (2018).

28. Ko, S. *et al.* GWAS of longitudinal trajectories at biobank scale. *The American Journal of Human Genetics* **109**, 433–445 (2022).

29. Wu, W. *et al.* Retrospective Association Analysis of Longitudinal Binary Traits Identifies Important Loci and Pathways in Cocaine Use. *Genetics* **213**, 1225–1236 (2019).

30. Xu, G. *et al.* Retrospective varying coefficient association analysis of longitudinal binary traits: Application to the identification of genetic loci associated with hypertension. *The Annals of Applied Statistics* **18**, 487–505 (2024).

31. Xu, H. *et al.* SPAGRM: effectively controlling for sample relatedness in large-scale genome-wide association studies of longitudinal traits. *Nat Commun* **16**, 1413 (2025).

32. Burrows, K. *et al.* A framework for conducting GWAS using repeated measures data with an application to childhood BMI. *Nat Commun* **15**, 10067 (2024).

33. Zou, T., Lan ,Wei, Wang ,Hansheng & and Tsai, C.-L. Covariance Regression Analysis. *Journal of the American Statistical Association* **112**, 266–281 (2017).

34. Frisch, R. & Waugh, F. V. Partial Time Regressions as Compared with Individual Trends. *Econometrica* **1**, 387–401 (1933).

35. Lovell, M. C. Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis. *Journal of the American Statistical Association* **58**, 993–1010 (1963).

36. Wang, W. & Yan, J. *Splines2: Regression Spline Functions and Classes*. (2024).

37. Wang, W. & Yan, J. Shape-Restricted Regression Splines with R Package splines2. *Journal of Data Science* **19**, 498–517 (2021).

38. Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. *Applied Longitudinal Analysis*. (John Wiley & Sons, Hoboken, New Jersey, 2011).

39. Gałecki, A. & Burzykowski, T. *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. (Springer, New York, NY, 2013). doi:10.1007/978-1-4614-3900-4.

40. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, giz082 (2019).

41. Basu, D. The Yule-Frisch-Waugh-Lovell Theorem for Linear Instrumental Variables Estimation. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.4514656 (2023).

42. Krishnakumar, J. Chapter 5 Time Invariant Variables and Panel Data Models: A Generalised Frisch–Waugh Theorem and its Implications. in *Contributions to Economic Analysis* (ed. Baltagi, B. H.) vol. 274 119–132 (Elsevier, 2006).

43. Ding, P. The Frisch–Waugh–Lovell theorem for standard errors. *Statistics & Probability Letters* **168**, 108945 (2021).

44. The MathWorks Inc. MATLAB version: 9.14.0 (R2023a). The MathWorks Inc. (2023).

45. Kristensen, K. & McGillycuddy, M. The AR(1) covariance structure. *Covariance structures with glmmTMB* https://cran.r-project.org/web/packages/glmmTMB/vignettes/covstruct.html#the-unstructured-covariance (2024).