ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Number 1

Assessing the Use of GEE Methods for Analyzing Continuous Outcomes from Family Studies: Strong Heart Family Study

Xi Chen [(1)], Ying Zhang [(2)], Amanda M. Fretts [(3)], Tauqeer Ali [(4)], Jason G. Umans [(5)], Richard B. Devereux [(6)], Elisa T. Lee [(7)], Shelley A. Cole [(8)], Yan D. Zhao [(9)]

(1) MD, PhD, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX. ORCID 0000-0002-5359-8902
(2) MD, PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0002-5359-8902
(3) PhD, MPH, Cardiovascular Health Research Unit, Department of Epidemiology. University of Washington, Seattle, WA. ORCID 0000-0002-5358-2345
(4) MD, MPH, PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0002-9176-879X
(5) MD, PhD, MedStar Health Research Institute, MedStar Health Research Institute, Hyattsville, MD, and Georgetown-Howard University, Washington D.C. ORCID 0000-0002-2746-3350
(6) MD, Weill Cornell Medicine, New York, NY. ORCID 0000-0002-8542-4982
(7) PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0003-1826-3602
(8) PhD, Texas Biomedical Research Institute, San Antonio, TX. ORCID 0000-0002-2651-0127
(9) PhD, Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0003-3448-0527

CORRESPONDING AUTHOR: Xi Chen, MD, PhD, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX. Email: xchen22@mdanderson.org.

SUMMARY

Background: Because of its convenience and robustness, the generalized estimating equations (GEE) method has been commonly used to fit marginal models of continuous outcomes in family studies. However, unbalanced family sizes and complex pedigree structures within each family may challenge the GEE method, which treats families as clusters with the same correlation structure. The appropriateness of using the GEE method to analyze continuous outcomes in family studies remains unclear. In this paper, we performed simulation studies to evaluate the performance of GEE in the analysis of family study data.
Methods: In simulation studies, we generated data from a linear mixed effects model with individual random effects. The random effects covariance matrix is specified as twice that of the pedigree matrix from the Strong Heart Family Study (SHFS) and other hypothetical pedigree structures. A Bayesian approach that utilizes the pedigree matrix was also conducted as a benchmark to compare with GEE methods with either independent or exchangeable correlation structures. Finally, analysis with a real data example was included.
Results: Our simulation results showed that GEE with independent correlation structure worked well for family data with continuous outcomes. Real data analysis revealed that all GEE and Bayesian approaches produced similar results.
Conclusion: GEE model performs well on continuous outcome in family studies, and it yields estimated coefficients similar to a Bayesian model, which takes genetic relationship into account. Overall, GEE is robust to misspecification of genetic relationships among family members.

Keywords: Bayesian; Generalized Estimating Equations; Kinship Matrix; Simulation; Strong Heart Family Study

DOI: 10.54103/2282-0930/20636
Accepted: 16th January 2023

INTRODUCTION

Generalized Estimating Equations (GEE) is a popular estimation approach used to fit marginal models on continuous outcomes in studies with repeated measurements or with clusters. Liang & Zeger first proposed the GEE method in *Biometrika* in 1986 [1]. By July 2021, their famous paper had achieved 19,086 citations. The popularity of GEE facilitates its incorporation in major statistical software, such as SAS, R, and STATA. A pivotal robustness property motivating widespread application of GEE is the high consistency and efficiency of the coefficient solution, no matter whether the working correlation structure is correctly specified.

Nevertheless, some previous studies indicated concerns about either the soundness of the theory or the proper use of GEE. For example, Crowder proposed that when the parameters used to calculate the working correlation matrix are uncertain in its definition, the asymptotic properties of the estimators can break down [2]. Mancl & Leroux revealed that the estimator yielded by the GEE model was fully efficient only for cluster-level covariates or covariates that are mean-balanced across clusters. In addition, the efficiency deceased as the variation in cluster sizes increased, and greater reductions occurred with higher between-cluster covariate variation [3]. Another study concluded that in GEE, misspecification of the correlation structure can be subject to a substantial loss in efficiency when covariates possess within-subject variability [4]. Furthermore, some critiques noted that GEE might not be the optimal model to use for data that are inherently unbalanced or for data with highly varied within-cluster correlation structures [5], although the systematic proof or simulations corresponding to this comment were not provided.

For family studies, such as the Strong Heart Family Study (SHFS) [6], data are correlated as a result of individuals being nested within each family. Depending on the scale of the study, the size of enrolled families can range from one to hundreds. In family studies, the kinship matrix is the statistical unit to store the information of relatedness among family members. Because of these wide-ranging family sizes, the kinship matrix can be complex and varies highly from one family to another. Such unbalanced family sizes and complex distribution of kinship matrix structures pose challenges in data analyses. Due to its convenience and popularity, GEE has been commonly used in data analysis in the Strong Heart Family Study [7-11]. However, without guidance from systematic simulation studies, it is unclear whether GEE is an appropriate approach with which to analyze family data.

When applying the GEE approach to family study data, there are a few concerns. First, to incorporate the kinship matrix defined among individuals, one random effect must be specified for each individual in the family study. Therefore, the total number of random effects is equal to the sample size. This contrasts with a typical GEE application in which random effects are defined at a cluster level. Moreover, GEE treated all families as clusters with an identical and simple correlation structure, which is far from the truth that the correlation structures among families are highly varied and can be very complex. Among the available GEE software packages, the correlation structures are predetermined and do not allow freedom in assigning distinct correlation structures across clusters.

In view of the aforementioned potential issues of applying the GEE method to analyze family data, we conducted simulation studies to evaluate the performance of the GEE method using a variety of simulation scenarios. Anticipating that the GEE method may not be appropriate to analyze family studies in certain scenarios, we also evaluated a Bayesian method proposed by Bae, Perls, & Sebastiani [12]. Their approach not only considers the within-cluster (a family) correlation by incorporating the kinship matrix in the model, but also avoids convergence issues due to the adoption of a singular value decomposition of the random effect covariance matrix. The Bayesian method was evaluated in the same way as the GEE method under the same simulation scenarios. At the end of this paper, we include analysis of SHFS data as an example of an application with which to compare GEE and Bayesian approaches.

The remainder of the paper is organized as follows. In Section 2, we described the statistical methods used in the study and briefly explained the derivation of the kinship matrix. In Section 3, we summarized the results from the simulation studies and analysis of the SHFS data. In Section 4, we highlighted the findings and discussed potential topics for future study.

METHODS

Conduct Simulation Using Linear Mixed Model (Conditional Model)

Linear mixed models (LMM) are commonly used to model continuous outcome variables obtained from correlated data. LMM include both fixed effects and random effects. In our study, to capture the kinship relationship among individuals with clusters, we specified random effects at individual levels.

Suppose we observe an outcome variable y in a sample with m families/clusters and a total sample size of n. Let n_i be the size for the ith family, $j=1,...,m$. The outcome for the jth individual in the ith family was generated by the model $y_{ij}=\mathbf{X_{ij}}\,\boldsymbol{\beta}+b_{ij}+\epsilon_{ij}$, where $\mathbf{X_{ij}}$ was the vector of covariates, $\boldsymbol{\beta}$ was the vector of fixed effects coefficients, b_{ij} was the individual-specific random effect that accounts for the additive polygenic effect, and ϵ_{ij} was the random error. For family i, we

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Number 1

stacked all of the random effects into a vector $\mathbf{b}_i=(b_{i1},...,b_{in_i})$, and we assumed $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2 \mathbf{A}_i)$, where σ_g^2 was the unknown genetic variance and $\mathbf{A}_i$ was the known correlation matrix, which was 2 times the kinship matrix $\mathbf{K}_i$.

Kinship matrix is a matrix consisting of kinship coefficients between any pair of individuals. The kinship coefficient K_{rs} for any two individuals r,s is the probability that genes selected randomly from r and s from the same autosomal locus are inherited from a common ancestor. Because the kinship sampling is done with replacement, when $r=s$ that is for the same person, $K_{rs}=1/2$. Table 1 lists kinship coefficients for several common types of relative pairs [13].

Table 1. Kinship coefficients for several common types of relative pairs

Relationship	Parent - Offspring	Half Siblings	Full Siblings	First Cousins	Uncle - Nephew
Kinship coefficient	1/4	1/8	1/4	1/16	1/8

In our simulations, two independent variables, age (continuous) and gender (binary), were included. The fixed effects of age (β_1) and gender (β_2) were set as $\beta_1=0.08$ and $\beta_2=-0.5$. The value of intercept was set as $\beta_0=1$. The value of genetic variance was set as $\sigma_g^2=1$. The random error was generated from a standard normal distribution $\epsilon_i \sim N(0,1)$. The values of the kinship matrix and independent variables were provided separately in two sets of simulations described below. In each of the simulation scenarios, we conducted 1,000 runs.

The first set of simulations was conducted using information obtained from the SHFS, a family-based prospective cohort study of cardiovascular diseases (CVD) and its risk factors among American Indians from 12 tribal communities in central Arizona, southwest Oklahoma, and North and South Dakota [6]. In our project, we adopted the baseline data of the SHFS. A total of 91 families with 2,764 individuals were included. Family sizes ranged from 1 to 113, with a median of 31, Q1 16 and Q3 39, with 78% of family sizes less than 40. The values of age and sex from the SHFS were adopted as the vector of covariates X_i. The SHFS kinship coefficients, which were derived from participant interviews and other lab work, were directly used to build up the kinship matrix $K_i, i=1,...,91$.

The second set of simulations was performed based on hypothetical families with selected kinship structures. The second data scenario was constructed to mimic a different kind of family data, in which kinship coefficients were not provided directly. Instead, the kinship matrix was derived by an R package kinship2, which requires variables of individual ID, individual's father ID and mother ID, and family ID to process the algorithm [14]. As an example, Figure 1 shows the data frame for a nuclear family (a), the pedigree plot (b), and the kinship matrix (c) calculated by the kinship2 package.

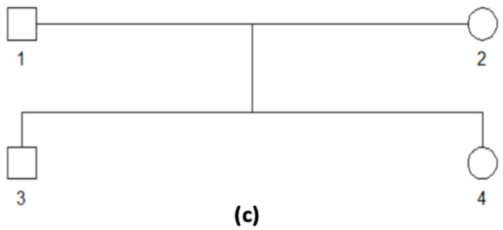
Figure 1. Data frame for a nuclear family (a), the pedigree plot (b), and the kinship matrix (c) calculated by the kinship2 package.

ID	Dad ID	Mom ID	Sex	Family ID
1	0	0	M	1
2	0	0	F	1
3	1	2	M	1
4	1	2	F	1

(a)

	1	2	3	4
1	0.50	0.00	0.25	0.25
2	0.00	0.50	0.25	0.25
3	0.25	0.25	0.50	0.25
4	0.25	0.25	0.25	0.50

(b)



(c)

Inspired by previous studies [12], we generated the corresponding variables of ID-series to create these family structures: (1) Singleton family: The family has only one member (same as independent data). (2) Nuclear family: The family structure is composed of a couple (father and mother) with two offspring. (3) Two-trios: This family structure is made up of first-, second-, and third-degree relatives, where two parent-offspring trios are related through a sibling pair in the parent generation. (4) Asymmetric family: This is an asymmetric and extended version of the second scenario, in which the first trio has only one offspring and the second trio has ten offspring.

In the second set of simulations, a family dataset with a total of 335 families and 1,020 individuals was generated. In each family, the gender of parents was defined as male as father and female as mother,

and the gender of children was randomly created by Bernoulli (0.5). The age of individuals was simulated by Uniform [a,b] for each generation of the family, with the boundaries of a and b set based on common logical order of parenthood, such that parents were older than the offspring and at least 25 years old.

For a linear outcome, the fixed effects coefficients in the conditional model and in the marginal model are equal mathematically. Since $y_{ij}=\mathbf{X_{ij}}\,\boldsymbol{\beta}+b_{ij}+\epsilon_{ij}$, and $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{A}_i)$, $\epsilon_{ij}\sim N(0,1)$, then the expectation of the outcome of the conditional model is $\mathbf{E}(y_{ij})=E(\mathbf{X_i}\,\boldsymbol{\beta}+b_i+\epsilon_i)=\mathbf{X_i}\,\boldsymbol{\beta}$, which is the expectation of the outcome variable in the marginal model. Thus, the assumed values of fixed effects in the simulated conditional model can be directly used as the true values of the fixed effects to evaluate the marginal model.

Generalized Estimating Equations (GEE) (Marginal Model)

The generalized estimating equations (GEE) method is the most common method to fit marginal models for longitudinal/clustered data. The GEE method uses an iterative algorithm to estimate regression coefficients and variance-covariance matrix. Standard errors for the estimates of regression coefficients are computed using a robust sandwich estimator. The "working" correlation structure in the synthesis of variance-covariance described the pattern of correlations within clusters. The independent correlation structure and exchangeable correlation structures were used in our study, as they are the top choices of analysis performed on family studies. Independent correlation structure assumes that any two of the individuals are independent in a cluster. Exchangeable correlation structure assumes that any two of the individuals share the same correlation. A previous study recommended that exchangeable correlation structure should be used for observations within a cluster, but without logical ordering [15]. The R package *geeM* was used to implement the GEE [16].

Bae's Bayesian Approach

To compare with GEE, we compared a novel Bayesian approach, in which the kinship matrix was incorporated to account for the within-family correlation [12]. For the frequentist approach, models with random effects are used to capture the correlation among individuals in family studies. However, due to the large family sizes, the high dimensionality of the random effects vector makes it difficult to converge [13, 14]. Bae et al. proposed to incorporate the singular value decomposition (SVD) in the Bayesian modeling approach in family studies to improve the non-convergence issue. The SVD was applied on the large covariance matrix of the random effect to "break down" the high dimensionality. In particular, for each family, $\mathbf{A}_i$ is decomposed by SVD, $\mathbf{A}_i=\mathbf{U}_i\,\mathbf{S}_i\,\mathbf{U}_i'$, where $\mathbf{U}_i$ is the matrix of eigenvectors and $\mathbf{S}_i$ is the diagonal matrix of eigenvalues. Define $\mathbf{b}_i=\mathbf{G}_i\mathbf{u}_i$, where $\mathbf{G}_i=\mathbf{U}_i\,\mathbf{S}_i^{1/2}$ and $\mathbf{u}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{I})$. We can show that $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{A}_i)$. Therefore, the random effect $\mathbf{b}_i$ was replaced by $\mathbf{G}_i\mathbf{u}_i$. For example, the model function for continuous outcome can be rewritten as $y_i=\mathbf{X}_i\,\boldsymbol{\beta}+\mathbf{G}_i\mathbf{u}_i+\boldsymbol{\epsilon}_i$. Bae et al. used non-informative priors for the parameters and provided BUGS code. In our study, JAGS and the R package *rjags* were used in the Bayesian approach, since JAGS shared the same coding language with BUGS.

EVALUATION OF GEE AND BAYESIAN APPROACH

The GEE and the Bayesian approach were both performed on simulated data. Relative bias and coverage probability from these two approaches were used to assess the point and interval estimation. Relative bias was calculated as the absolute bias divided by the true values. The coverage probability was calculated by the proportions that the true value of coefficient lies within 95% confidence intervals (or credible intervals) of coefficients generated in each simulation.

REAL DATA EXAMPLE

To further compare the GEE and Bayesian approaches, we performed analysis on real data obtained from SHFS. Suppose we aimed to investigate the factors that are related to systolic blood pressure: age, sex, body mass index (BMI), diabetes status, smoking, and alcohol consumption. Missing data were less than 1%, so a complete case data analysis was conducted. GEE (with independence and exchangeable correlation structures) and Bayesian approaches were both performed using the same software packages as were used in the simulation studies. Point estimates, standard error (standard deviation for the Bayesian model), and 95% confidence intervals (95% credible intervals for the Bayesian model) were compared.

RESULTS

Table 2 summarizes the results for the first set of simulations, which integrated the data from SHFS. Overall, all three models showed good performance. The relative biases were all close to zero and the coverage probabilities were all close to 95%. The GEE model with independence correlation structure performed slightly better than did the other two models. Table 3 summarizes metrics to evaluate models for the

second set of simulations in which hypothetical family structures were used. The results were similar to the those from the first set of simulations. There were no discernable differences in relative biases and coverage probabilities among the three models, and the GEE model with independence correlation structure seemed slightly better than the other two models.

Table 2. Comparison relative bias and coverage probability between GEE and Bayesian model approaches in simulated data based on kinship coefficients from SHFS.

Model	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Relative Bias			
Intercept	0.001	-0.004	-0.0004
Age	-0.0000007	0.0007	0.0005
Sex	-0.005	-0.003	-0.006
Coverage Probability			
Intercept	0.948	0.937	0.934
Age	0.943	0.941	0.954
Sex	0.952	0.941	0.937

[a]results with 1000 iterations, burn-in=100, chains=3, thin=2

Table 3. Comparison of bias, relative bias, and coverage probability between GEE and Bayesian models in simulated data based on kinship coefficients from a combination of singleton, nuclear, one-trio, two-trio, and three-trio families.

Model	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Relative Bias			
Intercept	0.00009	-0.0007	0.0002
Age	0.0004	0.0005	0.0004
Sex	-0.003	-0.003	-0.004
Coverage Probability			
Intercept	0.95	0.953	0.945
Age	0.942	0.942	0.939
Sex	0.945	0.945	0.945

[a]results with 1000 iterations, burn-in=100, chains=3, thin=2

The descriptive summary of the variables in the real data example is presented in Table 4. Participants in the study were middle aged, with mean age of 41 years, and generally overweight, with mean BMI 31kg/m^2. The majority of the participants were female (60%). The percentages of factors of interest were: diabetes (41%), current smoking (36%), and current drinking (58%).

Table 4. Descriptive summary of variables selected from SHFS for the analysis of real data

Variable	Mean	SD	Missing
Age	40.9	17.27	None
BMI	31.26	7.48	23
SBP	123	16.87	14
	Count	**Percent**	**Missing**
Sex (Female)	1649	59.70%	None
Diabetes	1115	40.60%	18
Current smoking	997	36.2%	10
Current drinking	1588	57.7%	12

Table 5 summarizes the point estimates, standard error (standard deviation for the Bayesian model), and the 95% confidence interval (credible interval for the Bayesian model) of the coefficients in each model. In general, the metrics were similar among the three models. The point estimates of the coefficients were very close. The Bayesian model tended to give smaller standard deviations than did the GEE models because GEE used robust sandwich estimation for the covariance matrix, while the Bayesian model made explicit distributional assumptions. For the 95% CI, the majority of the intervals were similar among the three models. There were disagreements on two covariates: diabetes and current drinking. The 95% CI of the two covariates covered zero for the GEE with exchangeable correlation structure, but were above zero for the other two models. The results were consistent with the fact that the p-values for the two covariables were around .05 for all three models.

Table 5. Summary of point estimates and standard error of model coefficients for analyses of SHFS data

	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Point Estimates			
Intercept	96.95	98.626	96.344
Age	0.41	0.41	0.416
Sex	-6.23	-6.328	-6.43
BMI	0.368	0.373	0.382
Diabetes	1.847	1.716	1.623
Current smoking	-0.113	-0.617	-0.334

Current drinking	1.487	2.267	2.204
Standard Error[b]			
Intercept	1.717	1.529	1.483
Age	0.023	0.023	0.018
Sex	0.681	0.666	0.549
BMI	0.05	0.044	0.039
Diabetes	0.89	0.886	0.637
Current smoking	0.719	0.698	0.589
Current drinking	0.786	0.734	0.61
95% CI[c]			
Intercept	(93.584, 100.316)	(93.629, 99.623)	(93.62, 99.31)
Age	(0.364, 0.456)	(0.364, 0.455)	(0.381, 0.449)
Sex	(-7.563, -4.895)	(-7.633, -5.024)	(-7.464, -5.406)
BMI	(0.27, 0.466)	(0.287, 0.459)	(0.31, 0.454)
Diabetes	(0.103, 3.59)	(-0.02, 3.452)	(0.382, 2.837)
Current smoking	(-1.523, 1.3)	(-1.985, 0.752)	(-1.525, 0.84)
Current drinking	(-0.053, 3.028)	(0.828, 3.705)	(0.99, 3.351)

[a] results with 2000 iterations, burn-in=100, chains=3, thin=5

[b] Standard deviation for the Bayesian Model

[c] Confidence Interval for GEE; Credible Interval for Bayesian Model

DISCUSSION

GEE serves as a handy tool for researchers to fit marginal models and make statistical inferences on clustered data, since this method can efficiently generate consistent estimates, regardless of the correct specification of within-cluster correlation structure. Data collected from a family study are clustered data. The unbalanced family sizes and the complex within-cluster relatedness may challenge the GEE performance. We evaluated the performance of GEE on simulated data with different types of family scenarios.

Our study is thus far the first to conduct systematic simulation studies to evaluate GEE in analysis of continuous outcomes in a family study. We simulated outcome data with covariates and kinship matrix from a real study, the SHFS, in which the kinship coefficients were generated based on meticulous interview and laboratory work. Furthermore, we included simulations with hypothetical family structures. We included a Bayesian model, which incorporated the kinship matrix in the modeling process, as a benchmark to compare with the GEE method. Results from the two sets of simulations indicated that both models work well, and there was no discernable difference between them. Moreover, the results of real data analyses revealed that the GEE and Bayesian models yielded similar estimates.

The performance of GEE on family data with continuous outcome was surprisingly good. When there is high correlation within responses, correct specification of correlation of responses potentially increases the efficiency. However, our integration of specific within-cluster correlation, the kinship matrix, did not bring much benefit in the Bayesian approach. When comparing the two GEE models, the independence correlation structure worked slightly better than the exchangeable correlation structure, which contradicts the fact that exchangeable correlation structure is recommended when there is no logical ordering for observations within a cluster [15]. It is possible that the simple structure improves the model fit efficiency. In conclusion, our results show that the GEE model performs well on continuous outcome in family studies, and it is robust to misspecification of genetic relationships among family members.

Our evaluation of GEE on family study focused on continuous outcomes, and our conclusion should not be simply applied to categorical or count outcomes. For continuous outcomes, the true values of regression coefficients in marginal models are equal to the values of fixed-effect regression coefficients in the conditional mixed effect models. This is because the linearity allows for the expectation of a continuous outcome to be calculated by the sum of expectation of each item in the model directly. However, categorical and count outcomes are typically modelled using a generalized linear model (marginal model) or a generalized linear mixed effects model (conditional model). Due to the nonlinearity of the link function, values of the regression coefficients in marginal models are no longer equal to those of in conditional models. Therefore, a direct comparison between the GEE approach and the Bayesian approach is not immediately available for categorical and count outcomes, and we leave this for future research.

ACKNOWLEDGEMENTS

The Strong Heart Study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, & 75N92019D00030. The study was previously supported by research grants: R01HL109315, R01HL109301, R01HL109284, R01HL109282,

and R01HL109319 and by cooperative agreements: U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Indian Health Service (IHS). Statistical analysis was partially supported by National Institutes of Health, National Institute of General Medical Sciences [Grant 2U54GM104938, PI Judith James].

REFERENCES

1. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13-22.
2. Crowder M. On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika. 1995;82(2):407-10.
3. Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. Biometrics. 1996:500-11.
4. Wang YG, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika. 2003;90(1):29-41.
5. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis: John Wiley & Sons; 2012.
6. North KE, Howard BV, Welty TK, Best LG, Lee ET, Yeh J, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. American journal of epidemiology. 2003;157(4):303-14.
7. Peng H, Zhu Y, Yeh F, Cole SA, Best LG, Lin J, et al. Impact of biological aging on arterial aging in American Indians: findings from the Strong Heart Family Study. Aging (Albany NY). 2016;8(8):1583.
8. Tinkelman NE, Spratlen MJ, Domingo-Relloso A, Tellez-Plaza M, Grau-Perez M, Francesconi KA, et al. Associations of maternal arsenic exposure with adult fasting glucose and insulin resistance in the Strong Heart Study and Strong Heart Family Study. Environment international. 2020;137:105531.
9. Jensen PN, Howard BV, Best LG, O'Leary M, Devereux RB, Cole SA, et al. Associations of diet soda and non-caloric artificial sweetener use with markers of glucose and insulin homeostasis and incident diabetes: the Strong Heart Family Study. European journal of clinical nutrition. 2020;74(2):322-7.
10. Grau-Perez M, Zhao J, Pierce B, Francesconi KA, Goessler W, Zhu Y, et al. Urinary metals and leukocyte telomere length in American Indian communities: the strong heart and the strong heart family study. Environmental Pollution. 2019;246:311-8.
11. Zhao Q, Zhu Y, Yeh F, Lin J, Lee ET, Cole SA, et al. Depressive symptoms are associated with leukocyte telomere length in American Indians: findings from the Strong Heart Family Study. Aging (Albany NY). 2016;8(11):2961.
12. 12. Bae HT, Perls TT, Sebastiani P. An efficient technique for Bayesian modeling of family data using the BUGS software. Front Genet. 2014;5:390.
13. Lange K. Mathematical and statistical methods for genetic analysis: Springer Science & Business Media; 2003.
14. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. Hum Hered. 2014;78(2):91-3.
15. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. The American Statistician. 1999;53(2):160-9.
16. McDaniel LS, Henderson NC, Rathouz PJ. Fast pure R implementation of GEE: application of the Matrix package. The R journal. 2013;5(1):181.