

Estimation of Parent-Sib Correlations for Quantitative Traits Using the Linear Mixed Regression Model: Applications to Arterial Blood Pressures Data Collected From Nuclear Families

Maha Al-Eid¹, Sarah AL-Gahtani² and Mohamed M. Shoukri^{3,*}

¹*Department of Biostatistics, Epidemiology and Scientific Computing, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia*

²*Department of Internal Medicine, King Fahd Medical City, Riyadh, Saudi Arabia*

³*Department of Epidemiology and Biostatistics, Schulich of Medicine and Dentistry University of Western Ontario, London, Ontario Canada*

Abstract: A fundamental question in quantitative genetics is whether observed variation in the phenotypic values of a particular trait is due to environmental or to biological factors. Proportion of variations attributed to genetic factors is known as heritability of the trait. Heritability is a concept that summarizes how much of the variation in a trait is due to variation in genetic factors. Often, this term is used in reference to the resemblance between parents and their offspring. In this context, high heritability implies a strong resemblance between parents and offspring with regard to a specific trait, while low heritability implies a low level of resemblance. While many applications measure the offspring resemblance to their parents using the mid-parental value of a quantitative trait of interest as an input parameter, others focus on estimating maternal and paternal heritability. In this paper we address the problem of estimating parental heritability using the nuclear family as a unit of analysis. We derive moment and maximum likelihood estimators of parental heritability, and test their equality using the likelihood ratio test, the delta method. We also use Fieller's interval on the ratio of parental heritability to address the question of bioequivalence. The methods are illustrated on published arterial blood pressures data collected from nuclear families.

Keywords: Genetic epidemiology, Familial correlations, Heritability, Linear Mixed normal models, Maximum Likelihood estimation, Estimating Ratio of parameters, Bootstrap Confidence interval.

INTRODUCTION

A heritable quantitative trait is a measurable phenotype that depends on the cumulative actions of many genes. These traits can differ among individuals, to produce a continuous distribution of phenotypes such as height, weight, blood pressures, high- and low-density cholesterol levels [1]. Quantitative traits also include molecular phenotypes such as gene expression levels [2].

Heritability analysis has been used for years to evaluate whether a given phenotype is influenced by genetic factors and how strong that influence is comparative to nongenetic risk factors. The general belief behind heritability analysis is that individuals who are more genetically related to each other should be more similar to each other for the phenotypes of interest [3].

The variation that exists in a quantitative trait can be divided into genetic and environmental components, and the genetic component can be additionally subdivided into additive dominance, and epistatic variances [4]. Estimation of the components of variance for a quantitative trait permits one to evaluate

both the degree to which genetics influences the trait and the trait's underlying genetic style [5]. The most relevant applications of heritability studies are those concerned with the heritability of components of Metabolic Syndrome (MS). The conclusions were, the genetic correlations seemed to vary under different conditions [6,7].

There is an increasing interest in the parental effect on their children with respect to Body Mass Index (BMI) and Blood Pressure (BP) levels, as the leading risk factors of heart diseases. Studying the heritability of components of metabolic syndrome has attracted the attention of genetic epidemiologists. A recent UK cohort study found that familial influence on BMI among middle-aged women appeared to be stronger from mothers than fathers [8]. A Canadian study reported that the prevalence of overweight and obesity among children and adults has risen in Canada. Studies suggest that parent obesity is a risk factor for overweight and obesity in children. This analysis examined associations between biological parent and child body mass index (BMI) in a nationally representative sample of Canadian children [9]. The study concluded that biological parent and child BMI were significantly correlated.

While the aim of the above studies was to establish parent offspring correlation, the issue of heritability has not been adequately addressed. For example, the sampling unit in the Canadian study [9] consisted of a

*Address correspondence to this author at the Department of Epidemiology and Biostatistics, Schulich of Medicine and Dentistry University of Western Ontario, London, Ontario Canada; Tel: (226-977-8651); E-mail: mmshoukr@uwo.ca, Shoukri.mohamed@gmail.com

parent and a child only, and ignored other siblings in the same family. Under this restricted sampling design, we cannot estimate the genetic components of the heritable trait of interest.

The fundamental aim of our paper is to establish statistical model-based approach to estimate heritability as defined in Laird and Lange [10]. Specifically, we will establish formal statistical methodologies to compare the paternal to maternal heritability, when the sampling unit is a cluster of nuclear families consisting of parents scores as well as their offspring scores for the same measured quantitative traits. The paper is structured as follows:

In section 2, we provide guidelines to the strategy of sampling nuclear families. We derive a formula for the optimal combination of the number of nuclear families and the optimal average sibship size needed to efficiently estimate parent heritability. In Section 3 we set up a regression model and identify the critical parameters that define parental heritability. We also derive moment estimators for heritability, the standard errors of the estimators, and construct confidence intervals on the ratio of parental heritability. We use the bootstrap methods to examine the distributional characteristics of the ratio estimator. In Section 4, and under the assumption of multivariate normality, we estimate the parent heritability and use the likelihood ratio test to evaluate the significance of the difference between maternal and paternal heritability. In Section 5, we illustrate the methodologies on published arterial blood pressures data. In Section 6 we present a general discussion

2. NUCLEAR FAMILY SAMPLING STRATEGY: ISSUES OF COST CONSTRAINTS

From the above studies, and because of the adopted study designs, it is impossible to account for the within siblings' correlation in estimating the correlation between parent and single offspring sampled at random from within the family. It is argued, as an example in the work by Kempthorne and Tandon [11] that one way to estimate parent heritability is done by using the regression of offspring trait on the parental trait, when the unit of study is the entire sibship. To estimate the required sample size (number of nuclear families and the average number of siblings within the family), we start with a model in which offspring scores is regressed on the one parent score (say the mother score).

Let Y_{ij} denote the score on the j^{th} offspring in the i^{th} family, and X_i , the score of the i^{th} parent, where $j=1, 2, \dots, n_i$, $i=1, 2, \dots, k$, and n_i is the number of offspring in the i^{th} family and k is the total number of

families. We assume that the regression of Y on X is given by:

$$Y_{ij} = \mu_y + \beta(X_i - \mu_x) + E_{ij} \quad (2.1)$$

where $\mu_y = E(Y_{ij})$, $\mu_x = E(X_i)$, β is the regression coefficient of Y on X , and E_{ij} is the deviation of the j^{th} offspring of the i^{th} parent. We further assume that

$$\text{Cov}(Y_{ij}, Y_{il}) = \begin{cases} \rho_e \sigma^2 & j \neq l \\ \sigma^2 & j = l \end{cases}$$

The most widely used estimator for β is given by the usual estimator:

$$b = \frac{\sum_{i=1}^k n_i (\bar{y}_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^k n_i (x_i - \bar{x})^2}$$

where $\bar{y} = \sum n_i \bar{y}_i / N$, $\bar{x} = \sum n_i x_i / N$, and $N = n_1 + n_2 + \dots + n_k$,

$$\bar{y}_i = \sum_j y_{ij} / n_i.$$

From [11] we have

$$V(b) = \frac{\sigma^2(1 - \rho_e) \sum n_i (1 + n_i \rho_e) (x_i - \bar{x})^2}{[\sum n_i (x_i - \bar{x})^2]^2}$$

To analyze clustered data, one must therefore model both the regression of Y on X and the within cluster dependence. If the responses are independent of each other, then ordinary least squares can be used, which produces regression estimators that are identical to the maximum likelihood in the case of normally distributed responses.

An important question that we need to answer at this stage is; what is the optimal number of sibships in order to efficiently estimate b . An important factor that needs to be considered is the cost of sampling nuclear families. To simplify our presentation, we shall assume that the variability among sibship sizes is relatively constant and that we can safely replace n_i by its average size, say n . In this case $V(b)$ reduces to:

$$V(b) = A(1 + n\rho_e)/nk \quad (2.2)$$

where $A = \sigma^2(1 - \rho_e)/S_{xx}$, and $S_{xx} = \sum_{i=1}^k (x_i - \bar{x})^2 / k$

Shoukri et al. [12] addressed the issue of obtaining the combinations (n, k) that minimize the variance of a specific estimator subject to cost constraints. In their attempt to construct a flexible cost function, they adhered to the general guidelines outlined by Flynn et al. [13]. First, one has to identify approximately the sampling costs and overhead costs. The sampling cost depends primarily on the size of the sample, and includes data collection costs, travel costs,

management and other staff costs. On the other hand, overhead costs remain fixed regardless of sample size, including, for example, the cost of setting the data collection form. Following Sukhatme *et al.* [14] it is assumed that the overall cost function is given as:

$$C = c_0 + kc_1 + nk c_2, \tag{2.3}$$

where c_0 is the fixed cost, c_1 the cost of recruiting an entire sibship, and c_2 is the cost of making a single observation within the sibship. Using the method of Lagrange multipliers [15], the objective function to be minimized G is given as

$$G = var(b) + \lambda(C - c_0 - kc_1 - nk c_2), \tag{2.4}$$

where $var(\hat{b})$ is given by equation (2.2), and λ is the Lagrange multiplier. The necessary conditions for the minimization of G are $\partial G/\partial n = 0$, $\partial G/\partial k = 0$ and $\partial G/\partial \lambda = 0$, with the sufficient condition for G to have constrained relative minimum given by a theorem in Rao [15]. Differentiating G with respect to n , k , and λ , equating to zero and solving for n , we obtain

$$n = \sqrt{R/\rho_e} \tag{2.5}$$

and

$$k = (C - c_0)/(c_1 + n c_2), \text{ where } R = \frac{c_1}{c_2}$$

For the sibship size to be meaningful, it is assumed that $\frac{R}{\rho_e} > 1$.

The interpretation of (2.5) is that we select larger sibships when R is large, that is when the cost of sampling an individual within a family is lower than sampling the entire sibship.

3. REGRESSION MODELS TO ESTIMATE MATERNAL AND PATERNAL HERITABILITY

Let the trait values of the i th family, of size n_i , be $(x_{im}, x_{if}, y_{i1}, y_{i2}, \dots, y_{in_i})$. The joint distribution of these random variables is characterized by the following parameters:

$$\left. \begin{aligned} E(y_{ij}) &= \mu_0 & \text{var}(y_{ij}) &= \sigma_0^2 \\ E(x_{im}) &= \mu_m & \text{var}(x_{im}) &= \sigma_m^2 \\ E(x_{if}) &= \mu_f & \text{var}(x_{if}) &= \sigma_f^2 \\ \text{cov}(x_{im}, y_{ij}) &= \rho_{m0} \sigma_m \sigma_0 & i &= 1, \dots, k \\ \text{cov}(x_{if}, y_{ij}) &= \rho_{f0} \sigma_f \sigma_0 & j &= 1, \dots, n_i \\ \text{cov}(y_{ij}, y_{il}) &= \rho_0 \sigma_0^2 & j &\neq l \\ \text{cov}(x_{im}, x_{if}) &= \gamma \sigma_m \sigma_f, \end{aligned} \right\} \tag{2.6}$$

Here x_{im} is the i th mother score, x_{if} is the i th father score and y_{ij} is the score of the j th offspring in the i th family. Moreover, we assume,

$$0 = \text{cov}(x_{im}, y_{ij}) = \text{cov}(x_{if}, y_{ij}) = \text{cov}(y_{ij}, y_{lk}) = \text{cov}(x_{im}, x_{lf}) \tag{2.7}$$

for all $i \neq l$. Following Mak and Ng [16] and Shoukri and Ward [17] we assume that model (2.6) has the following representation:

$$y_{ij} = \beta_0 + \beta_1 x_{im} + \beta_2 x_{if} + e_{ij} \tag{2.8}$$

where

$$\beta_0 = \mu_0 - \beta_1 \mu_m - \beta_2 \mu_f$$

and e_{ij} will have a mean zero and a covariance structure of intraclass correlation so that

$$\text{cov}(e_{ij}, e_{il}) = \begin{cases} \rho_e \sigma_e^2 & j \neq l \\ \sigma_e^2 & j = 1 \quad i = 1 \dots k \\ & j, l = 1 \dots n_i \end{cases}$$

and $\text{cov}(e_{ij}, e_{rl}) = 0, r \neq i$. It is also assumed that e_{ij} is not correlated with either x_{im} or x_{if} .

Using the definition of familial correlations provided by model (2.8) we can show that:

$$\sigma_0^2 = \sigma_0^2(\gamma) = \beta_1^2 \sigma_m^2 + \beta_2^2 \sigma_f^2 + 2\beta_1 \beta_2 \gamma \sigma_m \sigma_f + \sigma_e^2 \tag{2.9}$$

$$\rho_{m0} = \rho_{m0}(\gamma) = (\beta_1 \sigma_m + \gamma \beta_2 \sigma_f) / \sigma_0 \tag{2.10}$$

$$\rho_{f0} = \rho_{f0}(\gamma) = (\beta_2 \sigma_f + \gamma \beta_1 \sigma_m) / \sigma_0 \tag{2.11}$$

$$\rho_0 = \rho_0(\gamma) = (\beta_1^2 \sigma_m^2 + \beta_2^2 \sigma_f^2 + 2\gamma \beta_1 \beta_2 \sigma_m \sigma_f + \rho_e \sigma_e^2) / \sigma_0^2 \tag{2.12}$$

Conversely, equations (2.9-2.12) give:

$$\beta_1 = \frac{\sigma_0(\rho_{m0} - \gamma \rho_{f0})}{\sigma_m(1 - \gamma^2)} \tag{2.13}$$

$$\beta_2 = \frac{\sigma_0(\rho_{f0} - \gamma \rho_{m0})}{\sigma_f(1 - \gamma^2)} \tag{2.14}$$

$$\sigma_e^2 = \sigma_0^2 \left(1 - \frac{\rho_{m0}^2 + \rho_{f0}^2 - 2\gamma \rho_{m0} \rho_{f0}}{1 - \gamma^2} \right) \tag{2.15}$$

Equations (2.13-2.14) give ensemble estimators of the regression of offspring on their parental values, and equation (2.15) provide the estimator of the error term.

3.1. The Genetic Model

To describe the model that has both environmental and genetic components we assume a base population that is characterized by presence of additive genetic variance, and random environmental component. To simplify the estimation of heritability, we further assume that the interaction between genes and environment is negligible.

Therefore, the measured trait for the j th offspring in the i th family is such that

$$y_{ij} = \mu + 1/2(g_i - \mu_m) + 1/2(s_i - \mu_f) + \eta_{ij} + \varepsilon_{ij} \tag{3.1}$$

In equation (3.1), g_i is the breeding value of the i th mother, s_i is the breeding value of the i th father, and

η_{ij} is an individual deviation due to genetic segregation, and ε_{ij} is a random environmental deviation. Furthermore, we assume that:

$$E(g_i) = \mu_m, E(s_i) = \mu_f$$

$$\text{Var}(g_i) = \text{Var}(s_i) = \sigma_g^2$$

$$E(\eta_{ij}) = E(\varepsilon_{ij}) = 0$$

$$\text{Var}(\varepsilon_{ij}) = \sigma_0^2 - \sigma_g^2, \text{ with the restriction } \text{Var}(\eta_{ij}) = \frac{1}{2} \sigma_g^2$$

Note that the parental breeding values are not observable but may be estimated from the corresponding phenotype of the respective parent. This can be done under the assumptions:

$$g_i - \mu_m = h_m^2 (x_{im} - \mu_m) + D_{im},$$

$$\text{and} \tag{3.2}$$

$$s_i - \mu_f = h_f^2 (x_{if} - \mu_f) + D_{if}$$

Under the additivity of genetic effects, we can then use equation (3.2) to define the heritability in the narrow sense [10] as:

$h_m^2 = \sigma_g^2 / \sigma_m^2$, is the maternal heritability, and $h_f^2 = \sigma_g^2 / \sigma_f^2$ is the paternal heritability, as defined in the seminal work of Jacquard [18]. To obtain consistent estimators of (h_m^2, h_f^2) we assume that (x_{im}, x_{if}) are uncorrelated with the displacements (D_{im}, D_{if}).

Inserting (3.2) into (3.1) we get:

$$y_{ij} = \mu + \beta_1(x_{im} - \mu_m) + \beta_2(x_{if} - \mu_f) + e_{ij}$$

where

$$e_{ij} = (D_{im} + D_{if})/2 + \eta_{ij} + \varepsilon_{ij}$$

It can now be easily shown that

$$\sigma_g^2 = (h_m^2 \sigma_m^2 + h_f^2 \sigma_f^2) / 2$$

$$\sigma_e^2 = \sigma_0^2 - 1/4(h_m^4 \sigma_m^2 + h_f^4 \sigma_f^2) - 1/2 h_m^2 h_f^2 \gamma \sigma_m \sigma_f \tag{3.4}$$

$$\sigma_e^2 \rho_e = \text{cov}(e_{ij}, e_{il}) = \rho_0 \sigma_0^2 - 1/4(h_m^4 \sigma_m^2 + h_f^4 \sigma_f^2) - 1/2 h_m^2 h_f^2 \gamma \sigma_m \sigma_f$$

The first order approximation of the variances and the covariance of the estimators of the regression parameters are:

$$V_1 = \text{var}(\hat{\beta}_1) \approx \frac{\sigma_e^2}{\Delta} \left[\left\{ \sum_{i=1}^k \alpha_i(\rho_e) \right\} \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{if}^2 \right\} - \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{if} \right\}^2 \right]$$

$$V_2 = \text{var}(\hat{\beta}_2) \approx \frac{\sigma_e^2}{\Delta} \left[\left\{ \sum_{i=1}^k \alpha_i(\rho_e) \right\} \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{im}^2 \right\} - \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{im} \right\}^2 \right]$$

$$V_{12} = \text{cov}(\hat{\beta}_1, \hat{\beta}_2) \approx \frac{\sigma_e^2}{\Delta} \left[\left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{im} \right\} \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{if} \right\} - \left\{ \sum_{i=1}^k \alpha_i(\rho_e) \right\} \left\{ \sum_{i=1}^k \alpha_i(\rho_e) x_{im} x_{if} \right\} \right]$$

where

$$\Delta = \Sigma \alpha_i(\rho_e) \left[\left\{ \Sigma \alpha_i(\rho_e) x_{im}^2 \right\} \left\{ \Sigma \alpha_i(\rho_e) x_{if}^2 \right\} - \left\{ \Sigma \alpha_i(\rho_e) x_{im} x_{if} \right\}^2 \right] - \left\{ \Sigma \alpha_i(\rho_e) x_{im} \right\}^2 \left\{ \Sigma \alpha_i(\rho_e) x_{if}^2 \right\} - \left\{ 2 \Sigma \alpha_i(\rho_e) x_{im} \right\} \left\{ \Sigma \alpha_i(\rho_e) x_{if} \right\} \left\{ \Sigma \alpha_i(\rho_e) x_{im} x_{if} \right\} - \left\{ \Sigma \alpha_i(\rho_e) x_{if} \right\}^2 \left\{ \Sigma \alpha_i(\rho_e) x_{im}^2 \right\}.$$

and

$$\alpha_i = n_i \{ 1 + (n_i - 1) \rho_e \}^{-1}$$

4. TESTING THE HYPOTHESIS $H_0: h_m^2 = h_f^2$

This is equivalent to testing the equality of the regression parameters β_1, β_2 . This hypothesis will be tested using:

1. The delta method (DM) [19]
2. The Fieller's interval (FI) [20]
3. Bootstrap interval; (BI) [21]
4. The likelihood ratio test (LRT) [22]

We use the DM and FI to test the hypothesis $H_0: R = \beta_1 / \beta_2 = 1$, while we use the LRT to equivalently test $H_0: \beta_1 - \beta_2 = 0$.

4.1. The Delta Method

From [19] the asymptotic variance of the ratio of two random variables is given by:

$$\text{var}(\hat{R}) = \left(\frac{\partial \hat{R}}{\partial \beta_1} \right)^2 V_1 + \left(\frac{\partial \hat{R}}{\partial \beta_2} \right)^2 V_2 + 2V_{12} \left(\frac{\partial \hat{R}}{\partial \beta_1} \right) \left(\frac{\partial \hat{R}}{\partial \beta_2} \right) \tag{4.1}$$

Equation (4.1) gives the first order of approximation of the variance of the ratio estimator \hat{R} :

$$\text{var}(\hat{R}) = [V_1 + R^2 V_2 - 2RV_{12}] / \beta_2^2$$

We can therefore construct 95% confidence limits of the ratio R so that:

$$\text{Lower 95 \% confidence limit} = \hat{R} - 1.96 \sqrt{\text{var}(\hat{R})}$$

$$\text{Upper 95\% confidence limit} = \hat{R} + 1.96 \sqrt{\text{var}(\hat{R})}$$

It is well-known that the point estimator of the ratio parameter is biased. Using the DM, we have, to the first order of approximation:

$$\begin{aligned} \text{Bias}(\hat{R}) &= \frac{1}{2} \left[\left(\frac{\partial^2 \hat{R}}{\partial \beta_1^2} \right) V_1 + \left(\frac{\partial^2 \hat{R}}{\partial \beta_2^2} \right) V_2 + 2V_{12} \left(\frac{\partial^2 \hat{R}}{\partial \beta_1 \partial \beta_2} \right) \right] \\ &= 1/\beta_2^2 [R V_2 - V_{12}] \end{aligned} \tag{4.2}$$

4.2. Fieller's Interval

Let b_1 and b_2 denote respectively the moment estimators of β_1, β_2 .

$$\text{Let } U = b_1 - \delta_0 b_2$$

$$\text{Var}(U) = V_1 + \delta_0^2 V_2 - 2\delta_0 V_{12}$$

From [20], we assume that the rejection region of the hypothesis, with size α is such that:

$$\alpha = \Pr \left[\frac{(b_1 - \delta_0 b_2)^2}{V(u)} \geq Z_{\alpha/2}^2 \right]$$

From which,

$$\delta_0^2 [b_2^2 - V_2 Z_{\alpha/2}^2] - 2\delta_0 [b_1 b_2 - V_{12} Z_{\alpha/2}^2] + b_1^2 - b_1 Z_{\alpha/2}^2 \geq 0$$

$$\begin{aligned} \delta_0 &= \frac{2[b_1 b_2 - V_{12} Z_{\alpha/2}^2] \pm \sqrt{4(b_1 b_2 - V_{12} Z_{\alpha/2}^2)^2 - 4(b_2^2 - V_2 Z_{\alpha/2}^2)(b_1^2 - V_1 Z_{\alpha/2}^2)}}{2(b_2^2 - V_2 Z_{\alpha/2}^2)} \end{aligned}$$

Simplifying we get

$$\delta_0 = \frac{b_1 b_2 - V_{12} Z_{\alpha/2}^2 \pm [(b_1 b_2 - Z_{\alpha/2}^2 V_{12})^2 - (b_1^2 - Z_{\alpha/2}^2 V_1)(b_2^2 - Z_{\alpha/2}^2 V_2)]^{1/2}}{(b_2^2 - V_2 Z_{\alpha/2}^2)}$$

$$= A \pm B$$

where

$$A = \frac{b_1 b_2 - V_{12} Z_{\alpha/2}^2}{b_2^2 - V_2 Z_{\alpha/2}^2}$$

B

$$= \frac{[(b_1 b_2 - Z_{\alpha/2}^2 V_{12})^2 - (b_1^2 - Z_{\alpha/2}^2 V_1)(b_2^2 - Z_{\alpha/2}^2 V_2)]^{1/2}}{b_2^2 - V_2 Z_{\alpha/2}^2}$$

If the $100(1 - \alpha/2)\%$ Fieller confidence set for the ratio $R = \frac{\beta_1}{\beta_2}$ is finite, then it is given by $I = (A - B, A + B)$.

The confidence set is an interval if $b_2^2 - V_2 Z_{\alpha/2}^2 > 0$ and I is contained in $(\delta_0, \delta_0^{-1})$, that is if Fieller's confidence interval is included in the equivalence range. Usually $\delta_0 = 0.80 \Rightarrow \delta_0^{-1} = 1.25$. Hence the equivalence range is $(0.8, 1.25)$. These limits are shown in Figure 1.

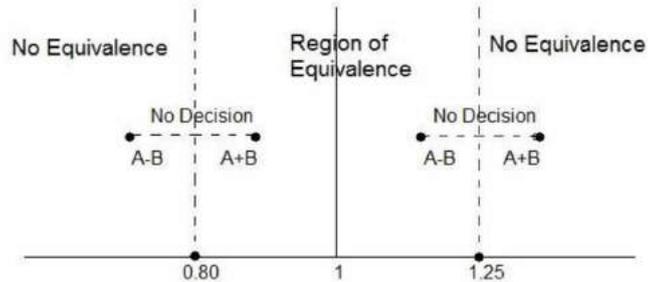


Figure 1: Limits of the Fieller's interval.

5. MAXIMUM LIKELIHOOD ESTIMATION: THE NORMAL LINEAR MIXED MODEL.

Let $X = (x_i; y_i)' \equiv (x_{im}, x_{if}; y_{i1}, \dots, y_{in_i})'$, $i = 1, 2, \dots, k$ be a random sample of k families. The likelihood function under model (2.2) will be given by

$$L = L(\Phi, x_i, y_i) = L_1(\Phi, x_i) L_2(\Phi, y_i | x_i) \tag{5.1}$$

Where

$$\Phi = (L_1; L_2)' = (\mu_m, \mu_f, \sigma_m^2, \sigma_f^2, \gamma; \beta_0, \beta_1, \beta_2, \rho_e, \sigma_e^2)'$$

$$\begin{aligned} L_2 &= (2\pi\sigma_e^2)^{-N/2} (1 - \rho_e)^{-(N-k)/2} \prod_{i=1}^k \{1 + (n_i - 1)\rho_e\}^{-1/2} \exp \left\{ -\frac{Q_2}{2\sigma_e^2(1 - \rho_e)} \right\}, \end{aligned} \tag{5.2}$$

With

$$Q_2 = SS_w + (1 - \rho_e) SS_B, N = \sum_i n_i$$

$$SS_w = \sum_i \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2, \bar{y}_i = \frac{1}{n_i} \sum_j y_{ij}$$

$$SS_B = \sum_i (\bar{y}_i - (\mu + \beta_1(x_{im} - \mu_m) + \beta_2(x_{if} - \mu_f)))^2$$

The relevant part of the global likelihood function is given by (5.2) and will be used to test the null hypothesis $H_0: \beta_1 = \beta_2 = \beta$ against all possible alternatives. We denote the maximized likelihood function under the alternative hypothesis by \hat{L}_a , and under the null hypothesis by \hat{L}_0 . Note that the part of the likelihood function that is affected by the restriction on the null hypothesis is SS_B . Under the null we write SS_B as shown in (5.3):

$$SS_B(0) = \sum_i^k (\bar{y}_i - \mu - \beta((x_{im} - \bar{x}_m) + (x_{if} - \bar{x}_f)))^2 \quad (5.3)$$

The null hypothesis is rejected when $H_{LR} = 2[\log(\hat{L}_a) - \log(\hat{L}_o)]$ exceeds $\chi_{1,\alpha}^2$, the upper cut-off value of a one degree of freedom chi-square distribution at α level of significance.

6. DATA ANALYSES

The methodologies presented thus fare are quite general. They are applicable to quantitative traits believed to be heritable and the study sample are nuclear families that include parents and their biological offspring. We illustrate the methodologies on data sets described below.

6.1. Example: Mial and Oldham’s Blood Pressures Family Data

The data used for illustration here are obtained from a survey that aimed at assessing the levels of similarity in systolic and diastolic blood pressure among family members living within 25 miles of Rhonda Fach Valley in South Wales. The data were published by Miall and Oldham [23]. Observations were made on parents and their offspring, with each observation consisting of

systolic and diastolic blood pressures measured to the nearest 5mm Hg. However, among 250 sampled families, only 204 contained information on brothers and sisters, whose age is above 18 years and live in the same household. Because of the impossibly low systolic blood pressure (15mm Hg) for one daughter, another family was omitted leaving 203 families for the analysis that had data on both parents and their offspring. The average number of siblings per family was 4 and the standard deviation was 1.92.

Before that data analysis we used R packages to produce pairwise correlations for parents and siblings blood pressure levels. Figure 2 shows the correlations and the histograms of each of the measured traits. This is quite informative since, in addition to the magnitude and the direction of correlations we can graphically detect the level of skewness of the measured traits. All the needed summary statistics are given in Table 1.

In Table 2 we summarize the results of both the delta and Fieller’s method. The delta method did not detect significance difference between the paternal and the maternal heritability for both types of blood pressures, since the interval includes unity. However,

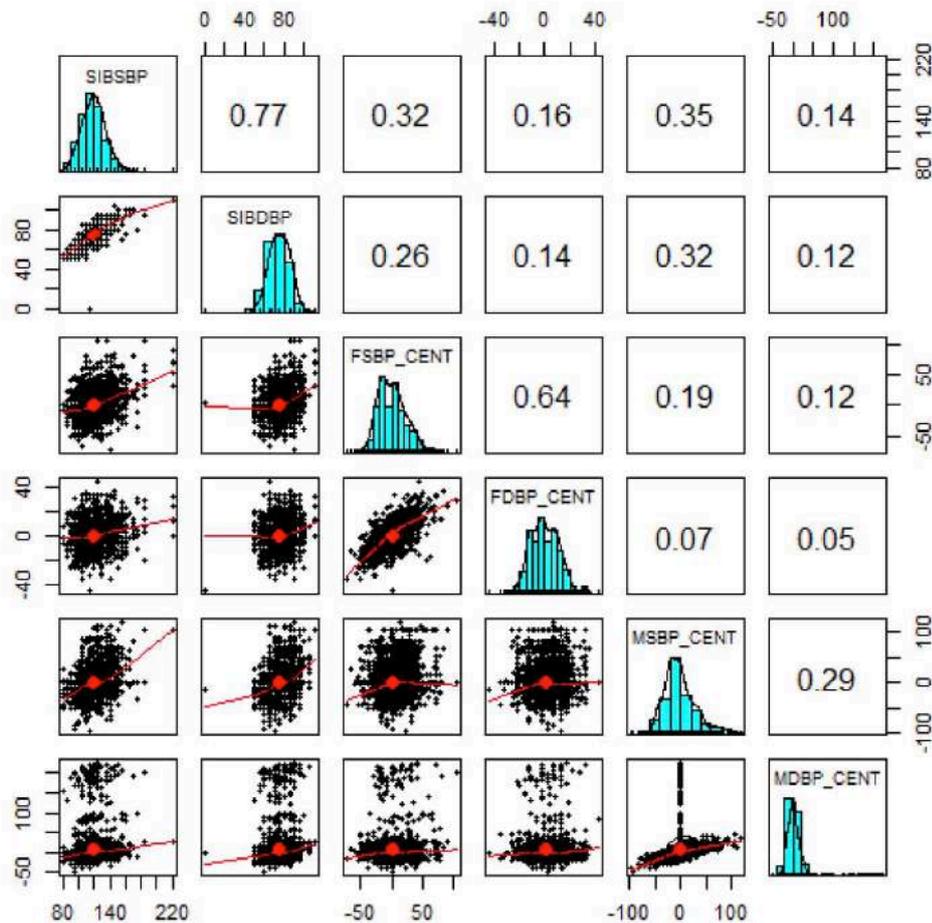


Figure 2: Pairwise correlations for familial blood pressures.

“This figure is quite informative as it displays the pair-wise correlations between pairs of variables measured at both offspring and parental levels”.

Table 1: Summary Statistics Including Correlations Based on the 2.9-2.15 Equations

Parameter estimates	Systolic Blood pressure	Diastolic blood pressure
σ_0	16.19	10.448
ρ_{mo}	0.367	0.272
ρ_{fo}	0.307	0.114
γ	0.163	0.029
σ_m	34.24	13.153
σ_f	22.954	11.558
Genetic parameters	$\beta_1 = 0.179 (0.025)$ $\beta_2 = 0.154 (.016)$ Cov (β_1, β_2) = -.00004	$\beta_1 = 0.096 (.034)$ $\beta_2 = 0.214 (.028)$ Cov (β_1, β_2) = -.0004
Environmental component	$\sigma_g^2 = 291.00$	$\sigma_g^2 = 45.195$
Sib-sib correlation	$\sigma_e^2 = 223.13$ $\rho_e = 0.566$	$\sigma_e^2 = 90.526$ $\rho_e = 0.333$

Table 2: Inference on the Ratio of Parental Heritability Using the Delta, Fieller's, and Bootstrap Resampling Technique

Method	Systolic Blood pressure (SBP)	Diastolic blood pressure (DBP)
1. Delta method	R=1.17 95% lower limit= .597 95% upper limit= 1.81	R= .448 95% lower limit= .245 85% upper limit= 2.688
2. Fieller's interval	95% lower limit= 0.606 95% upper limit= 1.268	95% lower limit= .125 95% upper limit= .917
Bias of ratio estimator	0.0185	0.0319
3. Bootstrap results	R=1.17(0.217) Q _{0.025} = 0.607 Q _{0.975} = 1.143 Bootstrap Bias = 0.026	R=0.448 (0.0896) Q _{0.025} = 0.066 Q _{0.975} = 1.223 Bootstrap Bias = 0.039

the Fieller's interval showed that for DBP the paternal heritability is significantly higher than the maternal heritability.

In order to verify the validity of the above methods (Delta and Fieller) we used the Bootstrap Resampling (BTR) to investigate the empirical characteristics of the ratio estimators. From Table 2 we see that the bootstrap method is in agreement with the other for the SBP, however based on the bootstrap interval, the hypotheses of equality of parental heritability was supported for the DBP, and this is in contrast to the conclusion by the Fieller's method. According to [21] one should trust the bootstrap results because it is completely non-parametric and does not require any distributional assumptions for the data. In Figures 3 and 5 we show the histogram of the bootstrap values of the ratios for SBP and DBP. As can be seen the distribution is quite skewed. This skewness is seen in Figures 4 and 6 as the plot of the empirical quantiles depart markedly from the quantiles of the normal distribution.

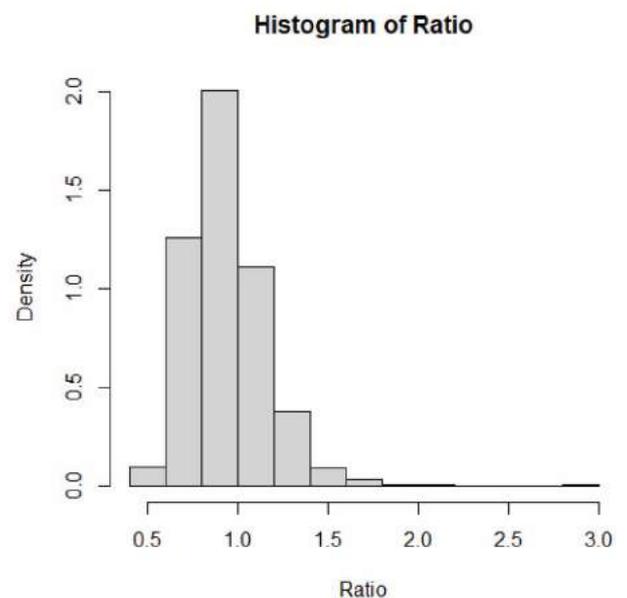


Figure 3: Histogram of the bootstrap distribution of Ratio for parental heritability for SBP.

Table 3 gives the results of the maximum likelihood estimation and the log-likelihood under the null and the

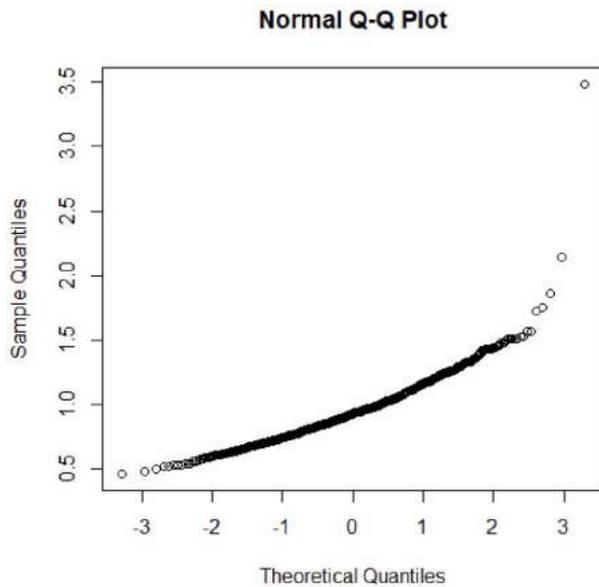


Figure 4: Graphical test of normality (Q-Q plot) of ratio of parental heritability for SBP.

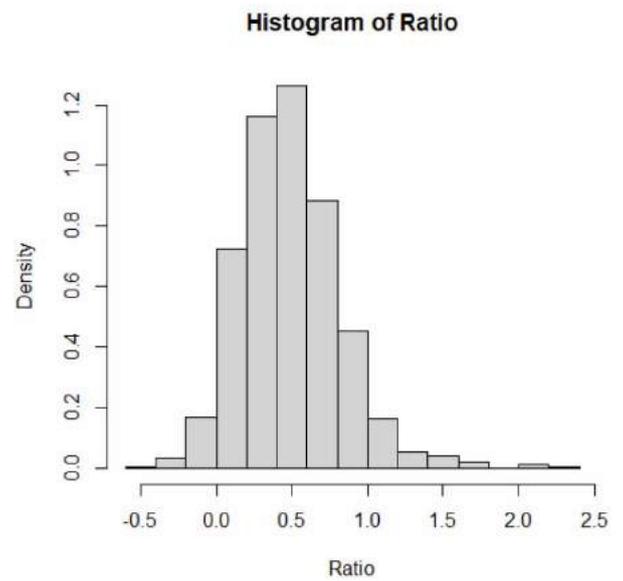


Figure 5: Histogram of the bootstrap distribution of the ratio of parental heritability for DBP.

alternative hypotheses. The likelihood-based approach has desirable asymptotic properties. The estimators are asymptotically normally distributed and asymptotically unbiased. For both blood pressure levels, the LRT detected significant difference between the maternal and the paternal heritability. The residual plots given in Figures 7 and 8 show that there are no outliers and they are almost normally distributed.

Table 3: Likelihood Ratio Test on the Null Hypothesis.

Mixed Likelihood analysis of SBP	Mixed Likelihood analysis of DBP
$\beta_1 = .01649 (.00099)$ $\beta_2 = 0.1269 (.0310)$ Log (L _a)= -11940.75 $\sigma_g^2 = 260.1869$ $\sigma_e^2 = 173.08$	$\beta_1 = .01123 (.0098)$ $\beta_2 = 0.2208 (.0498)$ Log (L _a)= -10980.22 $\sigma_g^2 = 247.885$ $\sigma_e^2 = 76.894$
Analysis under the null hypothesis:	Analysis under the null hypothesis:
$H_0: \beta_1 = \beta_2 = \beta$ $\beta = 0.1193 (.0275)$ Log (L ₀)= -11290.30 LRT= 1300 > 3.84	$H_0: \beta_1 = \beta_2 = \beta$ $\beta = 0.1133$ Log (L ₀)= -11934.18 LRT= 1908 > 3.84

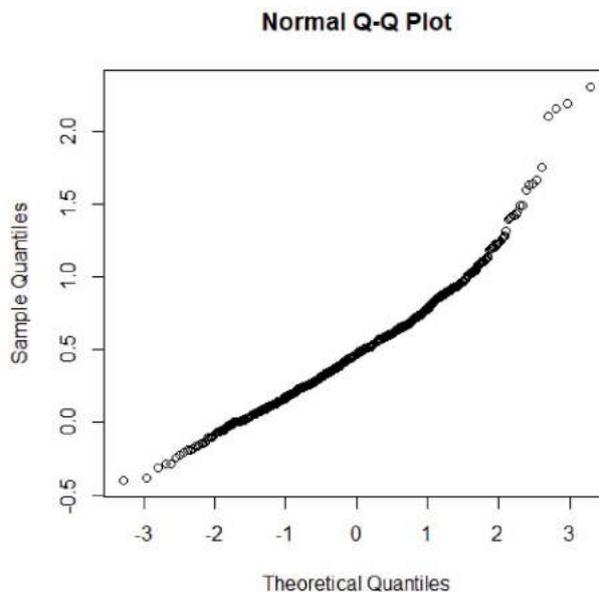


Figure 6: (Q-Q) plot for the ratio of Diastolic blood pressures.

We should not compare between the LRT and the other methods presented. This because the delta and the Fieller's methods test the significance of the departure of the ratios from unity. The LRT does not examine the ratio, rather it tests the significance of the difference from zero. Although, conceptually the two approaches are similar, the theoretical justifications are not the same.

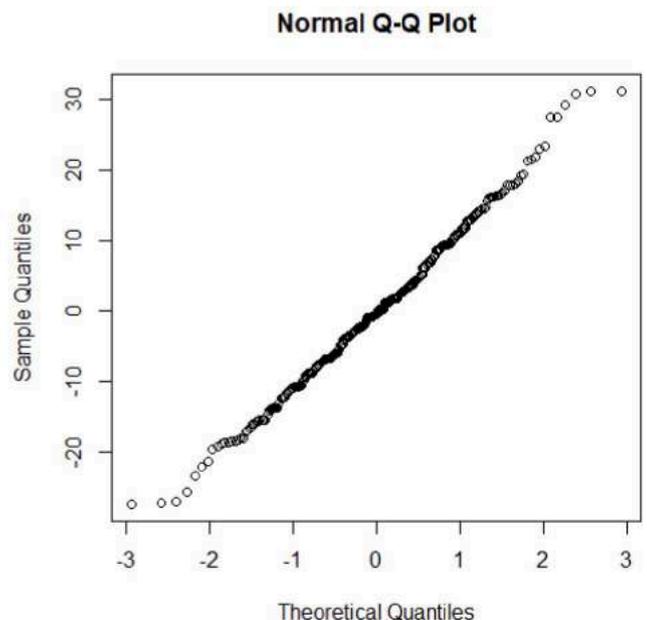


Figure 7: Residual plot under the linear mixed model for SBP data.

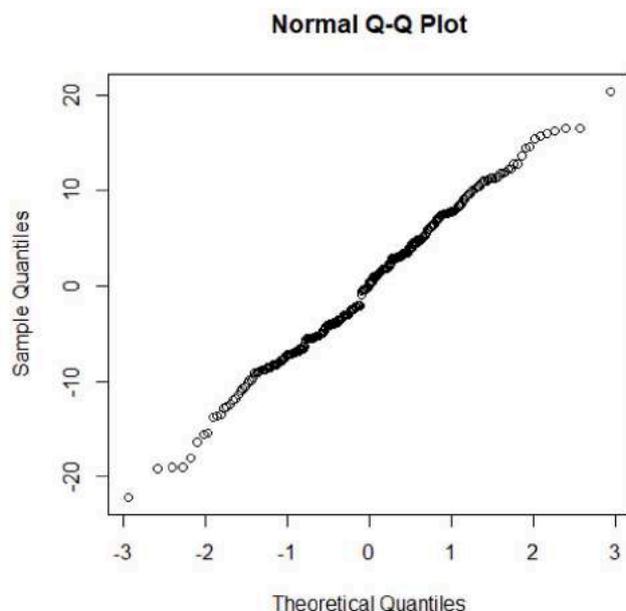


Figure 8: Residual plot under the linear mixed model for DBP data.

6. DISCUSSION

A recent study from China [25] on the familial aggregation of overweight (obesity) and high BP existed in rural areas and low-income families in China, respectively, suggesting both social and familial environments, alongside the impact of genetic, are important factors for non-communicable disease (NCD) risk factors. Furthermore, within two generations, considering offspring with highest BMI and BP were found to live with parents both having higher than normal BMI and BP, and strong father-offspring and mother-offspring correlations of BMI existed without substantial differences, both mother and father are indicated to play important roles in primary prevention strategies, and there might be great potential of family-based intervention against obesity. The salient fact is that “Non-communicable diseases (NCDs) are the leading cause of death and ill health and account for seven of ten deaths worldwide”. This study is quite important because the sampling units were nuclear families. However, the data analysis was descriptive and did not explicitly model the genetic components of variations.

A recent article appeared in the Lancet [26] outlined the WHO Sustainable Development Goal (SDG) target which is to reduce premature mortality from non-communicable diseases (NCDs) by a third by 2030 relative to 2015 levels. Among NCDs, heart disease is responsible for the highest risk of premature death in more than half of all countries for women, and more than three-quarters for men. Needless to say that elevated blood pressures levels are risk factors for heart diseases and examining the possibility of horizontal transmission from parents to their offspring

is an issue of public health at the family level. Estimation and inference procedures on the heritability of quantitative traits has been of interest to geneticists and genetic epidemiologists, medical geneticists, and genetic counselors. The questions of interests are; does a quantitative trait cluster with families; and does maternal heritability is higher or lower than parental heritability. We developed several inferential procedures to address both questions. All the suggested procedures, except the likelihood-based inference indicate that there are no significant differences in parental heritability for SBP, but differences exist for the DBP. The methods presented in this paper are quite general and are applicable to quantitative traits collected from nuclear families, and are applicable to molecular data as well. As we indicated in the introduction if a quantitative trait, such as blood pressure levels, BMI or waist-to-hip circumference are proven heritable, then the population risk of transmission of these traits from parents to their offspring may be reduced by genetic counselling through premarital screening strategy.

CONFLICT OF INTEREST

Non-Declared by all authors

APPENDIX: R CODE FOR BOOTSTRAPPING THE RATIO OF PARENTAL HERITABILITY.

```
names(data)
str(data)
head(data)
data=data.frame(na.omit(data))
attach(data)
summary(data)
B=1000
n=nrow(data)
n
Ratio=numeric(B)
##BOOTSTRAP ESTIMATE OF STANDARD ERROR
of Ratio
for (b in 1:B){
#randomly select the indices
i<- sample(1:n, size=n, replace=TRUE)
F<-data$FSBP_CENT[i]
M<-data$MSBP_CENT[i]
S<-data$SIBSBP[i]
f=sd(F)
m=sd(M)
```

```
gam=cor(F,M)
rfs=cor(F,S)
rms=cor(M,S)
original_ratio<-(f/m)*((rms-gam*rfs)/(rfs-gam*rms))
Ratio[b]=(f/m)*((rms-gam*rfs)/(rfs-gam*rms))
}
print(se.Ratio<-sd(Ratio))
hist(Ratio, prob=TRUE)
mean(Ratio)
bias_ratio<-mean(Ratio-original_ratio)
```

REFERENCES

- [1] Mackay TF. The genetic architecture of quantitative traits. *Annual Review of Genetics* 2001; 35(1): 303-339. <https://doi.org/10.1146/annurev.genet.35.102401.090633>
- [2] Mackay TF. Q&A: Genetic analysis of quantitative traits. *Journal of Biology* 2009; 8(3): 23. <https://doi.org/10.1186/jbiol133>
- [3] Pack AI, Keenan BT, Byrne EM, Gehrman PR. Genetics and genomic basis of sleep disorders in humans. In *Principles and practice of sleep medicine* 2017; pp. 322-339. Elsevier. <https://doi.org/10.1016/B978-0-323-24288-2.00031-3>
- [4] Cloninger CR, Van Eerdewegh P, Goate A, Edenberg HJ, Blangero J, Hesselbrock, Crowe R. Anxiety proneness linked to epistatic loci in genome scan of human personality traits. *American Journal of Medical Genetics* 1998; 81(4): 313-317. [https://doi.org/10.1002/\(SICI\)1096-8628\(19980710\)81:4<313::AID-AJMG7>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1096-8628(19980710)81:4<313::AID-AJMG7>3.0.CO;2-U)
- [5] Abney M, McPeck MS, Ober C. Broad and narrow heritability of quantitative traits in a founder population. *The American Journal of Human Genetics* 2001; 68(5): 1302-1307. <https://doi.org/10.1086/320112>
- [6] Must A, Spadano J, Coakley EH, *et al.* The disease burden associated with overweight and obesity. *JAMA* 1999; 282(16): 1523-1529. <https://doi.org/10.1001/jama.282.16.1523>
- [7] Gu D, He J, Duan X, *et al.* Body weight and mortality among men and women in China. *JAMA* 2006; 295(7): 776-783. <https://doi.org/10.1001/jama.295.7.776>
- [8] Johnson PC, Logue J, McConnachie A, *et al.* Intergenerational change and familial aggregation of body mass index. *Eur J Epidemiol* 2012; 27(1): 53-61. <https://doi.org/10.1007/s10654-011-9639-5>
- [9] Bushnik T, Garriguet D, Colley R. Parent-Child association in body weight status. *Health Reports. Statistics Canada, Catalogue no. 82-003-X.* June 2017; 28(6): pp. 12-19.
- [10] Laird N, Lange C. *The Fundamentals of Modern Statistical genetics.* Springer, New York 2011. <https://doi.org/10.1007/978-1-4419-7338-2>
- [11] Kempthorne O, Tandon O. The estimation of heritability by regression of offspring on parents. *Biometrics* 9: 90-100. <https://doi.org/10.2307/3001635>
- [12] Shoukri MM, Asyali MH, Walter SD. Issues of cost and efficiency in the design of reliability studies: Review and new results. *Statistical Methods in Medical Research* 2004. <https://doi.org/10.1191/0962280204sm365ra>
- [13] Flynn NT, Whitley E, Peters T. Recruitment strategy in a cluster randomized trial: cost implications. *Statistics in Medicine* 2002; 21: 397-405. <https://doi.org/10.1002/sim.1025>
- [14] Sukhatme PV, Sukhatme BV, Sukhatme S, Asok C. *Sampling theory of surveys with applications.* Iowa State University Press 1984.
- [15] Rao SS. *Optimization: theory and applications.* New Delhi: Wiley Eastern Limited 1984.
- [16] TK, Mak, Ng, KW. Analysis of familial data: Linear model approach. *Biometrika* 1981; 68: 457-461. <https://doi.org/10.1093/biomet/68.2.457>
- [17] Shoukri MM, Ward RH. Use of regression models to estimate genetic parameters and measures of familial resemblance in man. *The Journal of Royal Statistical Society, series C.* 1989; 38(3): 467-479. <https://doi.org/10.2307/2347734>
- [18] Jacquard A. Heritability: one word, three concepts. *Biometrics* 1983; 39: 465-477. <https://doi.org/10.2307/2531017>
- [19] Stuart, Allan, Ord, Keith. *Advanced Theory of statistics, Vol. 1 (Distribution theory), 6th edition.* J. Wiley, NY, USA 2010.
- [20] Fieller EC. "Some Problems in Interval Estimation." *Journal of the Royal Statistical Society, Series, (Methodological).* JSTOR 1954; 175-85. <https://doi.org/10.1111/j.2517-6161.1954.tb00159.x>
- [21] Davison AC, Hinkley, D. *Bootstrap methods and their applications,* Cambridge University Press 1997. <https://doi.org/10.1017/CBO9780511802843>
- [22] Stuart, Allan, Ord, Keith. *Advanced Theory of statistics, Vol. II (Statistical Inference), 5th edition.* Hafner publishing company 2009.
- [23] Miall WE, Oldham PO. A study of arterial blood pressure and its inheritance in a sample of the general population. *Clinical Science* 1955; 14: 459-487.
- [24] Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics.* Winter 1998; 24(3): 323-355. <https://doi.org/10.3102/10769986023004323>
- [25] Hu Y, He L, Wu Y, Ma G, Li L, Hu Y. Familial correlation and aggregation of body mass index and blood pressure in Chinese Han Population. *BMC Public Health* 2013; 13: 686. <https://doi.org/10.1186/1471-2458-13-686>
- [26] NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *Lancet* 2018; 392: 1072-88. [https://doi.org/10.1016/S0140-6736\(18\)31992-5](https://doi.org/10.1016/S0140-6736(18)31992-5)

Received on 02-10-2020

Accepted on 15-10-2020

Published on 03-11-2020

<https://doi.org/10.6000/1929-6029.2020.09.07>© 2020 Al-Eid *et al.*; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.