Variance components models for analysis of big family data of health outcomes in the Lifelines cohort study

Nino Demetrashvili¹,4,*, Nynke Smidt¹, Harold Snieder¹, Edwin R. Van den Heuvel¹,³ and Ernst C. Wit²,⁵

¹ Department of Epidemiology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands
² Bernoulli Institute, University of Groningen, 9747 AG Groningen, The Netherlands
³ Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
⁴ Department of Medical Statistics, National Center for Disease Control and Public Health, 0186 Tbilisi, Georgia, telephone: (+995)593727236
⁵ Institute of Computational Science, USI, 6900 Lugano, Switzerland
Abstract

Large multigenerational cohort studies offer powerful ways to study the hereditary effects on various health outcomes. However, accounting for complex kinship relations in big data structures can be methodologically challenging. The traditional kinship model is computationally infeasible when considering thousands of individuals. In this paper, we propose a computationally efficient alternative which employs fractional relatedness of family members through series of founding members.

The primary goal of this study is to investigate whether the effect of determinants on health outcome variable differs with and without accounting for family structure. We compare a fixed effects model without familial effects with several variance components models that account for heritability and shared environment structure.

Our secondary goal is to apply the fractional relatedness model in a realistic setting. Lifelines is a three generation cohort study investigating the biological, behavioral and environmental determinants of healthy ageing. We analyzed a sample of 89,353 participants from 32,452 reconstructed families. Our primary conclusion is that the effect of determinants on health outcome variables does not differ with and without accounting for family structure. However, accounting for family structure through fractional relatedness allows for estimating heritability in a computationally efficient way, showing some interesting differences between physical and mental quality of life heritability. We have shown through simulations that the proposed fractional relatedness model performs better than the standard kinship model not only in terms of computational time and convenience of fitting using standard functions in R, but also in terms of bias of heritability.
estimates and coverage.

KEYWORDS: Determinants and health outcome; Founders and non-founders; Fractional relatedness of family members; Genetic and environmental factors; Heritability and its confidence interval; Kinship model; Mental and physical health scores; Mixed effects models.
1 Introduction

Excessive weight, especially obesity, is a major public health concern in the western world. Around two third of the adult population in the United States and at least half of the population of many European countries are currently overweight or obese (Wang et al. 2008; Berghöfer et al. 2008; Flegal et al. 2010). In the Netherlands, the prevalence of overweight is 48.3% (Volksgezondheid 2012). 12.7% of the Dutch population is classified as obese. It is known that obesity – defined as a BMI of 30 or more – is a major risk factor for many chronic diseases, such as hypertension, stroke, coronary heart disease, diabetes, arthritis and overall mortality (Flegal et al. 2013; Prospective Studies Collaboration et al. 2009). Furthermore, increased BMI has also been shown to be associated with reduced physical Health Related Quality of Life (HRQoL) (Ul-Haq et al. 2012, 2013). However, evidence on the relationship between BMI and mental HRQoL is inconclusive. Some studies found that BMI is associated with poor mental health (Baumeister and Härter 2007; Ohayon 2007; Petry et al. 2008), whereas others did not find such relationship (Crisp and McGuiness 1976; Palinkas et al. 1996; Petry et al. 2008; Goldney et al. 2009; Ul-Haq et al. 2014) also found that the association between BMI and mental health among the Scottish adult population (n=37,272), was moderated by age and sex. In contrast to the above mentioned studies, Ul-Haq et al. (2014) used the full spectrum of BMI and adjusted for potential confounders and found that only young obese women (<45 years of age with BMI > 29.9 kg/m2) had significantly reduced mental health. Furthermore, being underweight was also associated with diminished mental health among women of all ages, but not men.
1.1 Background on Lifelines

Lifelines is a large population-based cohort study and biobank investigating the biological, behavioral and environmental determinants of healthy ageing among 167,729 inhabitants from the northern part of the Netherlands. The cohort profile of the Lifelines study has been extensively described in Scholtens et al. (2015). Summarizing, the participants baseline visit took place between December 2006 and December 2013. All general practitioners in the three northern provinces of the Netherlands were asked to invite their registered patients aged 25-49 years. All persons who consented to participate were asked to provide contact details to invite their family members (i.e., partner, parents, and children), resulting in a three generation study. In addition, participants could also register their participation via the Lifelines website. Lifelines adopted a multigenerational study design to disentangle the genetic, lifestyle and environmental contributions to the development of chronic diseases, study the between-generation similarities and identify the preclinical stages of ageing at an early age (Stolk et al., 2008). Baseline data were collected from 167,729 participants, of ages from 6 months to 93 years. Follow-up is planned for at least 30 years, with questionnaires administered every 1.5 years, and a physical examination scheduled every five years. The physical examinations, including anthropometry, lung function, blood pressure, electrocardiogram (ECG) and cognition tests are conducted at one of the Lifelines research sites. In addition, fasting blood and 24-hour urine samples are collected from all participants. A comprehensive questionnaire on history of (chronic) diseases, health related quality of life, lifestyle (physical activity, alcohol use, diet, smoking status), individual socioeconomic status (income and education level), psychosocial stress, work (profession, working hours), psychosocial characteristics and medication use is completed at home.
Lifelines is a facility that is open for all researchers. Information on application and data access procedure is summarized on [www.lifelines.net](http://www.lifelines.net). An overview of the available data is presented in the online Lifelines Data Catalogue.

### 1.2 Motivation and goal

Although the explicit familial structure of the data is an advantage in studying the genetic and environmental components of various health related outcomes, it is in fact a complicating factor in studying the effect of traditional epidemiological covariates. Given that participants are related (e.g. grandparent-grandchild, parent-child, sibling-sibling relations, etc.), genetic, shared environmental and health behavioral factors may confound the effects of BMI on physical and mental HRQoL.

The standard way to control for familial effects is through the kinship model ([Almasy et al.](Almasy et al. 1998)), which is a variance components model whereby genetic relatedness is modelled through a kinship covariance matrix. Despite the genetic plausibility of the model, it is computationally prohibitive for thousands of individuals and hundreds of families. This means that it cannot be used for the Lifelines study.

Our first objective is to study the effect of BMI on mental and physical components of HRQoL scores with and without accounting for relatedness in a family. Another objective is to determine a computationally efficient model that incorporates genetic and shared environment to assess the contribution of epidemiological determinants on health outcomes. Given the large sample size and inclusion of extended families in the study, Lifelines allows the identification of the effect of relatedness with higher precision than other studies that are often (much) smaller and limited to particular types of family relationships (e.g. [Pawitan et al.](2004); [Noh et al.](2006); [Yip et al.](2008); [Rabe-Hesketh et al.](2008); [Lichtenstein et al.](2008)).
The current Lifelines study includes extended families with up to 19 members.

Not only the variance components due to specific factors are of interests, but also the relative contribution of these variances in the total variance of the outcome. Related to this is the concept of intraclass correlation coefficient (ICC). The ICC represents the heritability coefficient in a narrow sense when applied to additive genetic models. The concept of heritability originates from Fisher (1919) and Wright (1920) and was formalized by Lush (1940). An extensive review on the concept and misconceptions of heritability is given by Visscher et al. (2008). In section 2.5 we provide the definitions of shared environmental, unique environmental and hereditary ICCs in the context of our models. We also briefly introduce the beta-approach (Demetrashvili et al. 2016) to construct confidence intervals for the ICC. The beta-approach has been successfully applied to construct confidence intervals for ratios of sums of variance components in linear and nonlinear mixed effects models (Demetrashvili and Van den Heuvel 2015). This approach will use the first and second moments of the ICC estimate in combination with a beta distribution for accurate confidence intervals.

In section 2 we provide the background on reconstruction of families and outcome measures. Then we describe various models and give a motivation of their use, including criteria for their selection. In section 3 we provide the analysis of familial confounding of BMI related physical and mental HRQoL in the Lifelines study. In section 4 we compare our fractional relatedness model with more traditional kinship models through simulations. We conclude the paper with an extensive discussion.
2 Methods

In this section we explain how we reconstructed extended families from the available local kinship relationships. The outcome of interest are mental and physical health, which are reconstructed from a RAND-36 questionnaire. Then we apply mixed effects models besides a fixed effect model to study the effect of BMI on mental and physical health.

2.1 Family reconstruction

For a large number of participants in the Lifelines study also their parents, partner and children are included in the study. Such information is relevant for disentangling the genetic, and behavioral and shared environmental variances. In biometric genetics, the coefficient of relatedness or genetic correlation for two individuals is defined as the expected proportion of genes of two individuals that are identical-by-descent (Sham 1998, p.208). A related concept in biometric genetics, which is used in this study, is that of a founder. Individuals without ancestors in the study are called founders, whereas others are called nonfounders (Almgren et al. 2003, p.10). Founders in our study population are assumed unrelated.

Considering the information provided by Lifelines participants we define a family as a group of related individuals sharing environmental and/or genetic factors. For example, health responses of mother and child may be similar due to both genetic similarity and shared behavior and environment, whereas the health responses of partners are related only due to the latter. Within the context of the Lifelines study we define the concept of an extended family as a connected graph of individuals either via parent-child or partner-partner relationships. An example of a family is given in Figure 1. Note that the sibling relationships in this graph are inferred from common parent relationships. Sibling information itself is not
recorded within Lifelines. Not for all reconstructed families in the Lifelines study do we have extensive information, as some members might not participate in the study.

Information on children from previous marriages is, in principle, also available. However, not all familial information is complete in Lifelines, since people may not be willing, for example, to identify ex-partners. Some of the information, however, can be reconstructed from the partial information provided by the participants. For example, if a child declares both parents and the parents declare this child, but do not declare each other as partners, we make a link between such parents as a couple. For family reconstruction we identify a set of related individuals through parent-child and/or partner-partner relationships and call this set a family. Once an individual is assigned to a particular family, no further reassignments take place. The model we propose in this study requires construction of the relatedness matrix. Relatedness between founders and non-founders is fractional, as explained in section 2.3. In Table 1 we outline the algorithm used for construction of fractional relatedness matrix.

### Table 1: Algorithm for construction of fractional relatedness matrix

1. For family $i$ determine the set of founders, i.e., individuals without parents.
2. Initialize a fractional relatedness matrix of size $n_i$ by $m_i$ as a matrix with zeros where $n_i$ is number of individuals and $m_i$ is number of founders in family $i$.
3. For each founder do the following: assign fractional relatedness equal to one to himself/herself and assign fractional relatedness equal to $(1/2)^g$ to offspring, where $g$ is a generational distance (e.g. $g=1$ for child, $g=2$ for grandchild, etc.).
2.2 Outcome measures

A sample consisting of 91,759 participants of the baseline Lifelines cohort study data release was available for analysis. HRQoL was measured using the Dutch version of the RAND-36 questionnaire (Van der Zee and Sanderman 1993; Van der Zee et al. 1996; Hays and Morales 2001). HRQoL refers to how health impacts on an individual’s ability to function and his or her perceived well-being in physical, mental and social domains of life. RAND-36 consists of 36 items measuring eight health concepts, i.e., physical functioning, role limitations caused by physical health problems, bodily pain, general health perceptions, role limitations caused by emotional problems, social functioning, emotional well-being and energy/fatigue. The first four reflect the physical health and the last four reflect the mental health of an individual. The scales of these eight concepts are combined into two summary measures of HRQoL, the physical component score (PCS) and the mental component score (MCS) using the scoring algorithm of Ware et al. (1994). PCS and MCS are between 0-100% and higher scores correspond to better quality of life.

2.3 Statistical models

We compare four models to examine whether the effect of BMI on HRQoL differs with and without accounting for family structure. These models are: $M_0$: multiple regression model; $M_1$: mixed effects model with random intercept for the family, capturing the environmental familial effect; $M_2$: mixed effects model with random slopes for founders within a family, capturing the genetic familial effect. The fourth model, $M_3$, is a combination of the last two models. Thus, $M_3$ is a mixed effects model with random intercept for family and random slopes for founders, capturing both environmental and genetic familial effects.

Assume that a total of $n$ individuals and $I$ families are included in the study.
Suppose for the $i$th family $n_i$ members have been observed ($i = 1, 2, \ldots, I$), such that $n = \sum_{i=1}^{I} n_i$. Multiple regression model $M_0$ of the HRQoL response $y_{ij}$ for the $j$th member of the $i$th family can be written as:

$$y_{ij} = x_{ij}^T \beta + \epsilon_{ij},$$  \hspace{1cm} (1)

where $x_{ij}$ is a $p \times 1$ vector for the $j$th individual in the $i$th family measured on $p$ covariates, including BMI and possible confounders, $\beta$ is a $p \times 1$ vector of coefficients, and residual errors $\epsilon_{ij}$ across all observations are assumed to be identically and independently distributed (iid) having a normal distribution with mean zero and variance $\sigma_R^2$, $\epsilon_{ij} \iid N(0, \sigma_R^2)$. The fixed effects $x_{ij}$ can be quantitative or dummy variables to represent categorical variables, so the effective number of covariates may be less or equal to $p$.

Note, observations $y_{ij}$ within the same family are most likely correlated, but the multiple regression model $M_0$ does not account for this. Model $M_1$ will account for this correlation by introducing a random intercept $u_i$ for every family:

$$y_{ij} = x_{ij}^T \beta + u_i + \epsilon_{ij}$$ \hspace{1cm} (2)

where $u_i$ is normally distributed with mean zero and variance $\sigma_u^2$, $u_i \iid N(0, \sigma_u^2)$; the definitions of $x_{ij}, \beta$ and $\epsilon_{ij}$ are the same as in (1).

Model $M_1$ does not disentangle the genetic and shared environmental variation. Since one of the goals is to estimate the variance contribution in MCS and PCS due to various factors, model $M_2$ will assume that the genetic correlation between family members is due to additive genetic effects of alleles (with no dominant and epistatic effects). By assuming all genetic information is in the founders, and consequently imposing the fractional relatedness effect between founders and
other members of the family, model $M_2$ introduces the random slopes $\upsilon_i$ for the set of founders $m_i$ in family $i$ and can be formulated as:

$$y_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{F}_{ij}^T \upsilon_i + \epsilon_{ij}$$  \hspace{1cm} (3)

where definitions of $\mathbf{x}_{ij}$, $\beta$ and $\epsilon_{ij}$ are the same as in (1) and $\mathbf{F}_{ij}$ is the $m_i \times 1$ vector of founders for individual $j$ in family $i$ where $F_{ijk}$ is a fractional relatedness of individual $j$ to founder $k$ in family $i$ with $\sum_{k=1}^{m_i} F_{ijk} = 1$; $\upsilon_i$ is the $m_i \times 1$ vector of random slopes for founders in family $i$ where $\upsilon_i = (\upsilon_{i1}, \ldots, \upsilon_{im_i})$, $\upsilon_{ik} \overset{iid}{\sim} N(0, \sigma^2_f)$. We assume independence between random terms.

An important computational advantage of model $M_2$ in comparison with traditional kinship model is employment of substantially smaller design matrix $\mathbf{F}_{ij}$. Namely, in $M_2$ the design matrix across all families is of size $m \times n$ where $m$ is the maximum number of founders among all families and $n$ is the total number of participants in all families. In classical kinship model the variance component matrix would be of size $n \times n$. Dimensionality reduction is particularly crucial when one analyzes large number of participants, such as in Lifelines.

An example of a family consisting of 13 members in the Lifelines study is shown in Figure 1: oval shapes refer to females and squares to males. The associated fractional relatedness matrix $\mathbf{F}$ is demonstrated below. The $M_2$ model assumes that the health outcome has a hereditary component. For example, the founders F1 and F2 both share 1/2 of their random hereditary effects with their child, member 3, and 1/4 with their grandchild, member 7.
Figure 1: Family example consisting of 13 members

\[
\begin{bmatrix}
F_1 & F_2 & F_{12} & F_{13} \\
1 & 1 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & 0 \\
3 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
4 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
5 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
6 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
7 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\
8 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\
9 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\
10 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\
11 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\
12 & 0 & 0 & 1 & 0 \\
13 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

Model $M_3$ is a combination of models $M_1$ and $M_2$, and can be written:

\[
y_{ij} = x_{ij}^T \beta + u_i + F_{ij}^T v_i + \epsilon_{ij}
\]  

(4)
where definitions and assumptions used in models $M_1$ and $M_2$ remain the same for this model. Variance components $\sigma^2_u$, $\sigma^2_f$, $\sigma^2_R$ are due to shared environmental, genetic and unique factors, respectively.

### 2.4 Inference and model selection

The overall significance of each covariate is tested using a conditional F-test. In this test, as in the usual F-test of covariates for regression models, the conditional estimate of the residual error variance is used. More details on the F-test are given in [Pinheiro and Bates (2009), §2.4.2](https://www.statisti.com/). Confidence intervals for marginal coefficients $\beta_l$ are constructed based on conditional t-tests. Each fixed effect coefficient can be tested marginally in the presence of other fixed effects in the model ([Pinheiro and Bates, 2009, pp. 92-96](https://www.statisti.com/)). The approximate 100 $(1 - \alpha)$ confidence limits on the $\beta_l$ are computed as:

$$
\hat{\beta}_l \pm t_{df_l}(1 - \alpha/2) \sqrt{\text{var}(\hat{\beta})_{ll}},
$$

(5)

where $\hat{\beta}_l$ is an estimate of $l$th fixed effect, $t_{df_l}(q)$ denotes the $q$th-quantile of a $t$-distribution with $df_l$ degrees of freedom and $\text{var}(\hat{\beta})_{ll}$ is an estimate of the variance of $\hat{\beta}_l$. Clearly, $\text{var}(\hat{\beta})$ is the variance-covariance matrix of the vector $\hat{\beta}$ of fixed effects estimates. More on the determination of the degrees of freedom for our models is in section 3.3.

To select the best model we used the Bayesian information criterion (BIC) ([Schwarz et al., 1978](https://www.statisti.com/)). The BIC for these models is defined as:

$$
BIC(M) = -2\ell_M(\theta|y) + df(M) \ln(n),
$$

(6)

where $\ell_M(\cdot)$ is the log-likelihood function for the estimated model $M$ with $\theta$ a vector of all parameters, $df(M)$ denotes the overall number of parameters in the
model, i.e., the regression and variance components parameters, and \( n \) is the total number of observations used to fit the model. The model with the smallest BIC is preferred.

### 2.5 Intraclass correlation coefficient

The models we defined above allow us to define the following three types of ICC: 1) the behavioral and shared environmental ICC (\( c^2 \)), as the proportion of total variance due to shared environmental components, 2) the hereditary ICC (\( h^2 \)), as the proportion of total variance due to the additive genetic component and 3) the unique environmental ICC (\( e^2 \)), as the proportion of total variance due to unique environmental components. We define these ICCs for model M₃, as follows:

\[
c^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_f^2 + \sigma_R^2} \quad h^2 = \frac{\sigma_f^2}{\sigma_u^2 + \sigma_f^2 + \sigma_R^2} \quad e^2 = \frac{\sigma_R^2}{\sigma_u^2 + \sigma_f^2 + \sigma_R^2}
\] (7)

For M₁ and M₂ models the \( c^2, h^2 \) and \( e^2 \) are deduced from formulae [7] by ignoring (setting to zero) those variance components that are not present in the model.

We outline the beta-approach for obtaining the CI for \( h^2 \), though this approach can be similarly applied to \( c^2 \) and \( e^2 \). The distribution of the estimator \( \hat{h}^2 \) is approximated with a beta distribution, \( \hat{h}^2 \sim \text{Beta}(a, b) \) with parameters \( a > 0 \) and \( b > 0 \). If \( \hat{h}^2 \) is an estimate of the mean and \( \hat{\tau}_{h^2}^2 \) is an estimate of the variance of \( \hat{h}^2 \), the method of moment estimates for \( a \) and \( b \) are given as:

\[
\hat{a} = \frac{\hat{h}^2(1 - \hat{h}^2) - \hat{\tau}_{h^2}^2}{\hat{\tau}_{h^2}^2} \quad \hat{b} = \frac{(1 - \hat{h}^2)[\hat{h}^2(1 - \hat{h}^2) - \hat{\tau}_{h^2}^2]}{\hat{\tau}_{h^2}^2}.
\] (8)

A first-order Taylor expansion is used to approximate \( \hat{\tau}_{h^2}^2 \), as shown in [Demetrashvili et al.](#).
The approximate 100\%(1 - \alpha) confidence interval on the \(h^2\) in (7) is then given by the lower and upper confidence limits as:

\[
LCL_{\hat{h}^2} = B^{-1}_{a,b}(\alpha/2),
\]
\[
UCL_{\hat{h}^2} = B^{-1}_{a,b}(1 - \alpha/2),
\]

with \(B^{-1}_{a,b}(q)\) being the \(q^{th}\)-quantile of the Beta(a, b) distribution. A detailed description of the beta-approach is given in Demetrashvili et al. (2016).

### 3 Lifelines analysis results

Analysis of the Lifelines data is conducted using R (R Core Team 2015), version 3.4.0. Mixed effect models have been fitted applying the \texttt{lme} function of the \texttt{nlme} package using maximum likelihood. Unlike for the traditional kinship model we were able to fit all models to thousands of subjects and families using the standard R function. All results below are presented with two-sided 95\% confidence intervals.

A sample of the baseline Lifelines cohort was used, consisting of 91,759 participants from which we constructed 32,531 families. The distribution of family sizes is summarized in Table 2. The largest family has 19 members. There are 253 singletons, 18,585 families have two members, etc. About 99\% of all reconstructed families consist of at least two members. Among all participants, 44\% were recruited via their general practitioner, 13\% via self-registration and the remaining 43\% were recruited as family members of the first two groups. The number of declared partners is 56,560, meaning that 62\% of all individuals in Lifelines currently have a partner. The number of declared fathers is 18,342 (20\%), whereas the number of declared mothers is 26,627 (29\%). 34\% of all individuals in Lifelines have at least one child. The maximum number of declared children is 7, which occurred
We omitted 2,406 (2.6%) observations for both MCS and PCS analyses. These observations were incomplete with respect to PCS scores, MCS scores or BMI. There were no missing values for sex or age. Finally, 89,353 observations were included in the analysis. The distribution of both outcomes, MCS and PCS, are slightly left-skewed. The median (25th, 75th percentile) MCS and PCS were 53.4 (48.9, 56.4) and 54.6 (50.5, 56.8), respectively. The distribution of family sizes for 89,353 participants are shown in Table 2. About 97% of the 32,452 reconstructed families with complete data consists of at least two family members. The maximum number of founders is 9.

Table 2: Counts of family sizes for the original set of 91,759 participants and the remaining 89,353 participants after removing incomplete records.

<table>
<thead>
<tr>
<th>family size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6-7</th>
<th>8-12</th>
<th>13-15</th>
<th>16-19</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>253</td>
<td>18,585</td>
<td>6,990</td>
<td>3,756</td>
<td>1,389</td>
<td>1,046</td>
<td>586</td>
<td>18</td>
<td>7</td>
<td>32,531</td>
</tr>
<tr>
<td>in analysis</td>
<td>883</td>
<td>18,446</td>
<td>6,863</td>
<td>3,542</td>
<td>1,303</td>
<td>962</td>
<td>433</td>
<td>14</td>
<td>6</td>
<td>32,452</td>
</tr>
</tbody>
</table>

3.1 Descriptive statistics of covariates

BMI is calculated by dividing a person’s weight measurement (in kilograms) by the square of their height (in meters) and subsequently categorized into six categories: underweight (< 18.5) kg/m², normal weight (18.5 – 24.9) kg/m², overweight (25.0 – 29.9) kg/m², class I obese (30.0 – 34.9) kg/m², class II obese (35.0 – 39.9) kg/m², class III obese (> 40) kg/m² [WHO 1995]. Out of 89,353 observations, 686 (0.7%) are underweight, 39,277 (44%) are normal, 35,824 (40.1%) are overweight, 10,494 (11.7%) are obese I, 2,306 (2.6%) are obese II, and 766 (0.9%) are obese III.

All subjects are 18 years and older, with an average age of 45. Out of 89,353 observations, 38,841 (44%) are men. Same proportions of males and females were
found, when all 91,759 observations were summarized.

3.2 Model selection, variance components and heritability

In the analysis of all models we used 89,353 observations with 32,452 constructed families. BMI is treated as a categorical variable with 6 levels. Age and sex are included in all models. Age is treated as a continuous variable. Sex is a categorical variable with male being a reference category. Model M0 is a multiple regression model with BMI, age and sex. Models M1, M2 and M3 are the variance components models. M1 models the environmental random component of the family. M2 models the genetic random components of the founders. M3 models both the random components of the family and founders.

We used BIC for model selection. BIC consistently selects the true model for large sample sizes (Claeskens and Hjort 2008) and tends to choose parsimonious models (i.e. models with few explanatory variables). Using the BIC model M1 provides the best fit for MCS, and M3 for PCS. Since the best model for MCS is modeling the shared environmental factors, the estimated ICC shown in Table 3 implies that approximately 12-14% variation in MCS is determined by shared environmental variation. Regarding the best model M3 for PCS, both the shared environmental as well as the genetic contribution are included, approximately 12-14% variation in PCS is determined by genetic variation and 3-4% by shared environmental variation.

3.3 Estimation of fixed effects

Besides accounting for the familial correlation structure, the aim of the study is to estimate the effect of BMI, age and sex on the HRQoL scores, PCS and MCS. The results of the conditional F-tests are presented in Table 4. The degrees of freedom for the tests of significance of slopes are 56,894, which are calculated by subtracting the number of families and the number of parameters of fixed effects.
Table 3: Estimates of variance components, ICC and its confidence interval for outcomes MCS and PCS; LCL and UCL stand for lower and upper confidence limits, respectively; Best model for outcomes MSC and PCS is in bold; \( \hat{c}^2 \) shows the proportion of behavioral and shared environmental variance in total phenotype variance and \( \hat{h}^2 \) shows the proportion of additive genetic variance in total phenotype variance.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Model</th>
<th>Variance component</th>
<th>ICC</th>
<th>ICC confidence</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estimator</td>
<td>Estimate</td>
<td>Estimate</td>
<td>LCL</td>
</tr>
<tr>
<td>MCS</td>
<td>M₀</td>
<td>( \hat{\sigma}^2_R )</td>
<td>64.842</td>
<td></td>
<td>626,443</td>
</tr>
<tr>
<td></td>
<td>M₁</td>
<td>( \hat{\sigma}^2_\mu )</td>
<td>8.438</td>
<td>( \hat{\sigma}^2_R )</td>
<td>56.551</td>
</tr>
<tr>
<td></td>
<td>M₂</td>
<td>( \hat{\sigma}^2_I )</td>
<td>5.855</td>
<td>( \hat{\sigma}^2_R )</td>
<td>60.206</td>
</tr>
<tr>
<td></td>
<td>M₃</td>
<td>( \hat{\sigma}^2_\mu )</td>
<td>8.438</td>
<td>( \hat{\sigma}^2_I )</td>
<td>2.11 × 10⁻⁶</td>
</tr>
<tr>
<td>PCS</td>
<td>M₀</td>
<td>( \hat{\sigma}^2_R )</td>
<td>45.546</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M₁</td>
<td>( \hat{\sigma}^2_\mu )</td>
<td>3.063</td>
<td>( \hat{\sigma}^2_I )</td>
<td>42.499</td>
</tr>
<tr>
<td></td>
<td>M₂</td>
<td>( \hat{\sigma}^2_\mu )</td>
<td>8.094</td>
<td>( \hat{\sigma}^2_I )</td>
<td>39.161</td>
</tr>
<tr>
<td></td>
<td>M₃</td>
<td>( \hat{\sigma}^2_\mu )</td>
<td>1.535</td>
<td>( \hat{\sigma}^2_I )</td>
<td>5.981</td>
</tr>
</tbody>
</table>

from the number of observations (i.e. 89,353-32,452-7). These degrees of freedom are also used in the t-test. The results for individual effects are presented in Figures 2 and 3, and in Table 5.

Figures 2 and 3 show the effect sizes of BMI (middle line) surrounded by confidence intervals (outer lines) of these effects. Plain lines are used for the selected models of MCS and PCS. Dashed lines are used for the other three models. Obviously, the effects of BMI on MCS and PCS of HRQoL match very closely across
### Table 4: Conditional F-tests for BMI, age and sex of selected models

<table>
<thead>
<tr>
<th>Model</th>
<th>Term</th>
<th>Numerator df</th>
<th>Denominator df</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS: $M_1$</td>
<td>BMI</td>
<td>5</td>
<td>56,894</td>
<td>72.6</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1</td>
<td>56,894</td>
<td>1,386.4</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>1</td>
<td>56,894</td>
<td>1,573.4</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>PCS: $M_3$</td>
<td>BMI</td>
<td>5</td>
<td>56,894</td>
<td>695</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1</td>
<td>56,894</td>
<td>412</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>1</td>
<td>56,894</td>
<td>2,460</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

### Table 5: Estimates of coefficients for BMI, age and sex of selected models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS: $M_1$</td>
<td>BMI (underweight)</td>
<td>47.502</td>
<td>0.313</td>
<td>46.889</td>
<td>48.114</td>
</tr>
<tr>
<td></td>
<td>BMI (normal)</td>
<td>48.890</td>
<td>0.099</td>
<td>48.696</td>
<td>49.083</td>
</tr>
<tr>
<td></td>
<td>BMI (overweight)</td>
<td>48.931</td>
<td>0.108</td>
<td>48.720</td>
<td>49.142</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 1)</td>
<td>48.790</td>
<td>0.129</td>
<td>48.538</td>
<td>49.042</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 2)</td>
<td>48.482</td>
<td>0.195</td>
<td>48.100</td>
<td>48.865</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 3)</td>
<td>48.364</td>
<td>0.307</td>
<td>47.763</td>
<td>48.965</td>
</tr>
<tr>
<td></td>
<td>Sex (female)</td>
<td>-1.891</td>
<td>0.053</td>
<td>-1.995</td>
<td>-1.788</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>0.078</td>
<td>0.002</td>
<td>0.074</td>
<td>0.081</td>
</tr>
<tr>
<td>PCS: $M_3$</td>
<td>BMI (underweight)</td>
<td>56.653</td>
<td>0.259</td>
<td>56.145</td>
<td>57.161</td>
</tr>
<tr>
<td></td>
<td>BMI (normal)</td>
<td>57.491</td>
<td>0.082</td>
<td>57.331</td>
<td>57.650</td>
</tr>
<tr>
<td></td>
<td>BMI (overweight)</td>
<td>56.731</td>
<td>0.090</td>
<td>56.556</td>
<td>56.906</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 1)</td>
<td>54.927</td>
<td>0.107</td>
<td>54.716</td>
<td>55.137</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 2)</td>
<td>52.978</td>
<td>0.164</td>
<td>52.656</td>
<td>53.299</td>
</tr>
<tr>
<td></td>
<td>BMI (obese 3)</td>
<td>50.373</td>
<td>0.258</td>
<td>49.868</td>
<td>50.879</td>
</tr>
<tr>
<td></td>
<td>Sex (female)</td>
<td>-1.011</td>
<td>0.046</td>
<td>-1.100</td>
<td>-0.921</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>-0.081</td>
<td>0.002</td>
<td>-0.084</td>
<td>-0.078</td>
</tr>
</tbody>
</table>

four models (lines overlap) and this is true for all categories of BMI. Confidence intervals also match very closely. This implies that the BMI effects do not change with and without accounting for relatedness in a family.

We see inverted parabolic shape of MCS across increasing BMI categories. A somewhat different shape is observed for PCS across increasing BMI categories.
Interestingly, the MCS slightly increases for overweight people compared to the normal category. It also shows a dramatic drop for underweight individuals. The PCS decreases for all categories of BMI compared to the normal category and the decreasing trend of PCS is increasingly steeper with increasing BMI. The confidence intervals of all BMI categories for PCS are beyond the confidence limits of the normal category, meaning that all BMI categories have substantially different PCS than the normal category.

![BMI effects for MCS](image1)

![BMI effects for PCS](image2)

Figure 2: BMI effects for MCS surrounded by confidence intervals

Figure 3: BMI effects for PCS surrounded by confidence intervals

In Table 5 we present the coefficients for fixed effects with measures of uncertainty, namely standard error, lower and upper confidence limits. Results show that each additional year of age is associated with a 0.078 unit increase (0.1% of maximum observed value) in MCS and decrease by about the same amount in PCS, on average, holding BMI and sex constant. Women have lower MCS, by approximately 1.891 units (2.6% of maximum observed value) and lower PCS, by approximately 1.011 units (1.4%), on average, holding BMI and age constant.
4 Simulation study: design and results

We conducted a simulation study to compare the commonly used variance component model with kinship matrix [Almasy et al. 1998] to our suggested model containing a reduced matrix of founders. The kinship model is computationally demanding and therefore we had to simulate relatively small sample sizes. We generated data using the model (10) shown below:

\[
y_{ij} = x_{ij}^T \beta + u_i + k_{ij} \sigma^2_f + \epsilon_{ij}
\]

where definitions of \(x_{ij}, \beta, u_i, \epsilon_{ij}, \sigma^2_u, \sigma^2_R\) are the same as in model (2), (3), (4). \(\sigma^2_f\) is the variance due to genetics and \(k_{ij}\) is the kinship effect for the \(j\)th member of the \(i\)th family. \(k_{ij}\) is generated from the multivariate normal distribution with mean zero and variance-covariance matrix \(\Omega\). The \(\Omega\) consists of coefficients of pairwise relationships formed in the following way: in the first degree of relationship (parent-child, siblings) the coefficient of relationship is 1/2, in the second degree of relationship (grandparent/grandchild, half-sibling, avuncular) the coefficient of relationship is 1/4 and similarly for other degrees, as shown in Table 1 of [Almasy et al. 2010]. Technically, \(\Omega\) is a block-diagonal matrix with coefficients of relationships among family members on the diagonal blocks and zeros (the coefficients of relationship between families) on the off-diagonal blocks.

Simulation parameters were selected from the results of best fitted model for Lifelines PCS data (see parameters in PCS: M of Table 5) and equal to: \(\beta_{\text{underweight}} = 56.5, \beta_{\text{normal}} = 57.5, \beta_{\text{overweight}} = 56.7, \beta_{\text{obese1}} = 54.9, \beta_{\text{obese2}} = 53.0, \beta_{\text{obese3}} = 50.2, \beta_{\text{age}} = -0.08, \beta_{\text{sex}} = -1.0, \sigma^2_u = 1.5, \sigma^2_f = 6.0, \sigma^2_R = 40.0.\)
Variable for age was generated from the normal distribution with mean 45 and standard deviation 14. Variable for sex was generated from Bernoulli distribution with probability 0.44 (for men). Number of observations for BMI categories “underweight”, “normal”, “overweight”, “obese 1”, “obese 2”, “obese 3” were generated from the multinomial distribution with probabilities 0.007, 0.44, 0.401, 0.117, 0.026 and 0.009 respectively. These probabilities were calculated from the Lifelines data. Then the BMI was generated from a normal distribution based on the number of observations from the multinomial distribution and the following mean and standard deviation parameters: 17.7, 0.77 for “underweight”, 23.0, 1.5 for “normal”, 27.0, 1.4 for “overweight”, 32.0, 1.4 for “obese 1”, 37.0, 1.4 for “obese 2”, 43.0, 3.0 for “obese 3”. For family sizes we set 2, 3, 4, 5, and 6 members, each size repeated 2, 4, 10, 20, and 38 times respectively. Even though family sizes were repeated, in fact all families were different in their composition (e.g. for family size of 2 we constructed of parent-child and partner-partner families; for size of 3 we constructed families of parent-parent-child, parent-child-grandchild, parent-child-child and parent-child-partner of child). In total, 74 families were constructed containing 384 numbers of observations. The PCS outcomes were generated using model \(10\). Afterwards we fitted both models (kinship, \(M_3\)) and compared the biases of heritability estimates and coverage probabilities of 95% confidence intervals of heritabilities. Heritability in model \(10\) is equivalent to hereditary ICC \((h^2)\) and computed as shown in \(7\). Confidence intervals for heritability were calculated using the beta-approach, as described in Section \(2.5\). We conducted 100 simulations. The kinship model was fitted using the lmekin function of the coxme (\cite{therneau2015coxme} package and model \(M_3\) using lme function of nlme (\cite{pinheiro2017nlme} package in R. Results of simulation studies are summarized in Table \(6\).

Results from setting 3 are visualized in Figure \(4\). The kinship model results in
Table 6: Comparison of heritability parameters and computational time between kinship (lmekin function) and \( M_3 \) (lme function) models

<table>
<thead>
<tr>
<th>Setting</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of families</td>
<td>19</td>
<td>37</td>
<td>74</td>
</tr>
<tr>
<td>Total number of observation</td>
<td>96</td>
<td>192</td>
<td>384</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( M_3 ) model (seconds)</th>
<th>( M_3 ) model</th>
<th>( M_3 ) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time per simulation</td>
<td>0.11</td>
<td>0.12</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Time per simulation in kinship model (seconds)</td>
<td>0.92</td>
<td>2.69</td>
</tr>
<tr>
<td>True heritability</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Bias of heritability (^1) in ( M_3 ) model</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>Bias of heritability (^1) in kinship model</td>
<td>0.19</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Coverage of 95% confidence intervals of heritability in ( M_3 ) model</td>
<td>0.78</td>
<td>0.91</td>
<td>0.98</td>
</tr>
<tr>
<td>Coverage of 95% confidence intervals of heritability in kinship model</td>
<td>0.66</td>
<td>0.88</td>
<td>0.86</td>
</tr>
</tbody>
</table>

larger bias (0.18) in comparison with the one (0.09) in \( M_3 \) model. The \( M_3 \) model results in larger variation than the kinship model on average, as shown in Figure 4. Subsequently, the kinship model demonstrates substantial undercoverage (0.86) while the \( M_3 \) shows some overcoverage (0.98) for a two-sided 95% confidence interval.

We compared the computational time needed for fitting the kinship and \( M_3 \) models. We varied the number of families and, correspondingly, the total number of observations. We examined 3 settings, each with 100 simulations. Settings and results are shown in Table 6. \( M_3 \) is 8 times faster than the kinship model (setting 1). As the number of families quadruples (from setting 1 to 3), the computational time increases 2.5 times for \( M_3 \) model (lme function) and 15.2 times for kinship model (lmekin function). Thus, there is an exponential increase in computational time for the kinship model as the number of observations increases while there is much slower increase (roughly linear) in computational time for the \( M_3 \) model.

\(^1\)heritability is equivalent to hereditary ICC (\( h^2 \)) and computed using formula (7).
M3 model estimates heritability better than kinship model

Figure 4: Overlapping histograms for comparison of estimated heritabilities across M3 and kinship models in setting 3: vertical bar (0.13) shows true heritability, light grey histogram (left) and smoothed plain line show distribution of estimated heritabilities in M3 model, dark grey histogram (right) and smoothed dashed line show distribution of estimated heritabilities in kinship model.

5 Discussion

In this study the primary goal was to answer whether relatedness in a family must be accounted for when estimating the effects of risk factors of interest in large family-based cohort studies. The answer to this question is of particular importance for researchers analyzing the Lifelines data and data from other large...
cohort studies of families. From our study it is clear that the effects of BMI on MCS and PCS of HRQoL scores do not change when accounting for family structure. This conclusion confirms theoretical considerations within longitudinal data analysis given by Diggle et al. (2002, chapter 1). The authors state that when the focus of the study is on modeling the dependence between the response and explanatory variable, then the nature of correlation among responses is unimportant if there is a large number of families relative to the number of individuals per family. Our Lifelines data clearly satisfies this criterion.

McArdle et al. (2007) conducted a simulation study with the objective to compare the performance of association analysis of family based designs that account for and ignore family structure in assessment of the phenotype-genotype association. They concluded that effect size estimates and power are not significantly affected by ignoring family structure. Though, type 1 error rates increase when family structure is ignored, and the magnitude of the increase depends on trait heritability and pedigree configuration. Induced type 1 error is directly related to diminished standard errors (and narrow confidence intervals) leading to liberal inference about regression coefficients, i.e., falsely claiming significance when there is none. In our analysis of both PCS and MCS, we saw that ignoring the correlation (or family structure) lead instead to larger standard errors of the regression coefficients (although the increase was very small), thereby leading to conservative inference about the covariate effects. Therefore, the standard errors of regression parameters can be larger or smaller when ignoring family structure, and therefore may lead not only to liberal inference and inflated type I errors, but also to conservative inference and deflated type I errors. Increase or decrease of the standard errors depends on (i) the relationship between the family structure and the covariates of interest and (ii) the family structure effect size on the outcome.
We compared the Lifelines analysis results on association between BMI and HRQoL with similar results in the literature. Increased BMI has been shown to be associated with reduced physical HRQoL (Ul-Haq et al. 2012, 2013), however, evidence on the relationship between BMI and mental HRQoL were antagonistic. In the Lifelines study we see reduced mental HRQoL for all categories of BMI in comparison with the overweight category. This conclusion matches with the conclusion of Ul-Haq et al. (2013) from meta-analysis study. Furthermore, an inverted U-shape of mental HRQoL across increasing BMI categories is seen in both our Lifelines study (Figure 2) and that of Scottish study conducted by Ul-Haq et al. (2012). Similarly to Ul-Haq et al. (2013) we see that increasing BMI is associated with impaired HRQoL in Lifelines. In addition, our study reveals a shape of association between BMI and physical HRQoL. There is an inverted J-shape negative association in PCS across increasing BMI categories (Figure 3).

In this work we studied real data and did not assume any a priori family inheritance structure. The main strength of our study is the use of Lifelines data which makes it possible to estimate the relatedness in more complex families than other studies can. Consequently, our results are practically more relevant. We learned that with or without accounting for family structure, the effect of determinants on health outcomes do not significantly change. Nevertheless, the ability to incorporate family structure into our model in a computationally efficient way allows one to disentangle the genetic, shared behavioral and environmental variances. Furthermore, our proposed model allows for estimating hereditary, behavioral and shared environmental, and unique environmental ICCs of both HRQoL outcomes in computationally efficient way through fractional relatedness of founders and non-founders.

The kinship model as implemented in the SOLAR (Almasy et al. 1998) software
is unable to handle the analysis of the Lifelines data with over 89,000 individuals, although we did not study exhaustively all methods, such as generalized estimating equations \cite{Liang and Zeger 1986} that could have been used to implement the kinship model. We overcame the computational infeasibility with kinship model by introducing and fitting a variance component model with fractional relatedness using standard functions in R.

We have made the assumption in our study that founders are unrelated, but Lifelines may not have complete information. For example, siblings information itself is not recorded within Lifelines, and therefore siblings might be modeled as two founders while they do have a common ancestor and are related. It would be interesting to examine whether modeling just a subset of founders would impact the variance component parameter estimates.

Furthermore, MCS and PCS may be genetically correlated, meaning that there is genetic overlap between these two traits (i.e., the same set of genes may regulate these traits). If interest lies in separation of genetic and environmental contributions simultaneously in MCS and PCS, then bivariate models could be used, similarly to the way that others \cite{Yip et al. 2008, Lichtenstein et al. 2009} modeled schizophrenia and bipolar disorder using multivariate generalized linear mixed models.

In summary, the proposed model offers to solve the computational issues involved in modeling family structure when thousands of families are analyzed, and subsequently to fit the model accurately using standard functions of R.

\section{Acknowledgments}

The authors wish to acknowledge the services of the Lifelines cohort study, the contributing research centers delivering data to Lifelines, and all the study participants.
participants. The authors declare no conflict of interest.

7 Financial Support

During the first six months this work was supported by the University of Groningen, University Medical Center Groningen as part of PhD thesis and later it was supported by STAR travel grant from the University of Groningen. For the remaining work this research received no specific grant from any funding agency, commercial or not-for-profit sectors.

References


Ware, J.E., Kosinski, M., and Keller, S.D. (1994) SF-36 physical and mental summary scales: a user’s manual. The Health Institute, Boston, MA.

