

Statistical Issues in the Evaluation of Clustering of Metabolic Syndrome in Spousal Pairs

Shoukri MM^{1,2*}, Collison K¹ and Al-Mohanna F^{1,2}

¹Department of Cell Biology, King Faisal Specialist Hospital and Research Center, Saudi Arabia

²College of Medicine, Al-Faisal University, Saudi Arabia

Abstract

Background: The metabolic syndrome is intimately linked hypertension, impaired glucose tolerance, dyslipidemia, and abdominal obesity and is associated with an increased risk of total and cardiovascular mortality in adults. Genetics as well as environmental influences have been implicated in obesity and several cardiovascular risk factors. Because Family is one of the most important factors affecting metabolic risk factors, studying co-aggregation of the components of the syndrome among family members, and in particular spousal pairs, is of interest to genetic epidemiologists and community health researchers.

Methods: Based on the clinical definition of the syndrome, we introduce three statistical models to estimate the prime parameter of interest which measure the degree of clustering of the disease among spousal pairs. Since the focus in this paper is on the methodological approach to estimate the between pairs clustering parameters, we shall use Monte-Carlo simulated data for demonstration purposes, with values of the input parameters for each component taken from a well-known Korean study. We develop two models, the Bivariate Truncated Poisson Model (BTPM), which models the counts, and the Bivariate Dirichlet Multinomial Model (BDMM), which models the frequency of counts, and discuss the relative merits of each model. The two models are qualitatively different but quantitatively interrelated. Since the clinical definition of the metabolic syndrome requires that at least three of its components, co-exist within a subject, we show that adhering to this definition requires certain specifications that should be satisfied in any of the adopted models.

We estimated the clustering parameters under the specified models. A comparison between the models was based on the internal consistency of each model. What we mean by that is the degree of closeness of the estimated distribution to the observed data. The BDMM fitted the data much closer than the BTPM.

Interpretation: In a sample of randomly selected spousal pairs; and according to the clinical definition, the number of components of the metabolic syndrome can in an individual be 0, 1, 2, 3, 4, or 5. Estimation of the clustering parameter of the counts is equivalent to the estimation of the intraclass correlation coefficient (ICCC) between pairs. Assessing the goodness of fit of the proposed models, it is more statistically sound to estimate the degree of clustering of the components of the syndrome in spousal pairs under the BDMM.

Keywords: Statistical issues; Evaluation; Metabolic syndrome; Pairs

Introduction

The metabolic syndrome is the co-aggregation of hypertension, impaired glucose tolerance, dyslipidemia, and abdominal obesity and is associated with an increased risk of total and cardiovascular mortality in adults [1,2]. Genetics as well as environmental influences have been implicated in obesity and several cardiovascular risk factors [3,4]. Family is one of the most important factors affecting metabolic risk factors in children, in that family displays an interaction between genetic and shared environmental factors [5,6]. Recent research showed that childhood and adolescent overweight has been increasing in Asian countries due to urbanization and economic development. For example, over the past 10 years, the rates of overweight among Korean children and adolescents aged 5-20 years have doubled [7], which may ultimately cause an increase in adverse cardiovascular outcomes. Globally, the prevalence of the metabolic syndrome is high among obese children and adolescents, and increases with increasing obesity [8].

Most of the research on metabolic syndrome is related to description by Reaven [9] of syndrome X or 'the insulin resistance syndrome'. In 1988, Reaven et al. [10] used the term syndrome X to refer to the tendency of glucose intolerance, hypertension, low high density lipoprotein (HDL) cholesterol and raised triglycerides, and hyperinsulinaemia to occur in the same individual.

Broadly, the research on this highly prevalent condition followed two lines of work: one based on the epidemiological studies whose main concern has been to identify risk factors for cardiovascular disease; and the other based on clinical and experimental studies concerning the pathogenesis of diabetes and atherosclerosis. We focus in this paper on statistical and epidemiological aspects of metabolic syndrome.

Given the importance of the metabolic syndrome as a public health problem, estimation of its population prevalence according to the most recent definitions poses challenges. Within the frame work of clustering or familial aggregation of traits we focus on the issue when the sampling strategy is cross sectional cluster sampling with two units within each cluster; husband and wife, or spousal pairs. The major concern with our research is to formulate model from which measures

*Corresponding author: Shoukri MM, Department of Cell Biology, King Faisal Specialist Hospital and Research Center, 11211 Riyadh, Saudi Arabia, Tel: +966-509491454; E-mail: shoukri@kfshrc.edu.sa

Received May 29, 2015; Accepted June 10, 2015; Published June 18, 2015

Citation: Shoukri MM, Collison K, Al-Mohanna F (2015) Statistical Issues in the Evaluation of Clustering of Metabolic Syndrome in Spousal Pairs. J Biom Biostat 6: 233. doi:10.4172/2155-6180.1000233

Copyright: © 2015 Shoukri MM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of spousal concordance are the prime parameters of interest.

As was noted by Laird and Lange [11], “The general concepts used in aggregation and heritability analysis are widely accepted as useful measures of the degree to which traits are inherited; most researchers would not undertake genetic analysis without evidence of aggregation or heritability of the trait.” One should note that familial aggregation of a trait is a necessary but not sufficient condition for inferring the importance of genetic susceptibility, since environmental and cultural influences can also play a role in familial clustering and excess familial risk. For quantitative traits, the biometrical approach introduced by Morton [12], Rao et al. [13], and Morton and McLean [14] to evaluate the degree of resemblance among family members has relied on the well-developed multivariate normal theory. However in assessing the degree of family resemblance, clinical epidemiologists often prefer to report the disease status of individuals on a categorical scale. Therefore analytic approaches established under the multivariate normal model are not useful. It should also be emphasized that questions regarding to familial aggregation of traits can be effectively addressed under appropriate sampling designs. Extended nuclear families, sib-ships based sampling designs, and twins studies are examples of designs that are often adopted by genetic epidemiologists to answer questions regarding heredity and environmental impacts on traits of interest.

Spousal resemblance [15] or concordance may be due to shared environment, common behaviors, and also positive assortative mating, that is, the tendency of individuals to choose a spouse with similar characteristics. If concordance was mainly due to a cohabitation effect, then it should increase with increasing time shared by spouses. Differential effects of cohabitation and assortative mating are not mutually exclusive, and both should be considered for a correct interpretation of spousal resemblance. Therefore, spousal resemblance is a subject of interest that can be studied under the spousal-pairs design.

The paper is structured as follows. In Section 2 we estimate the clustering parameter and construct an overall estimate pooled from each component. In Section 3 we discuss the estimation based on the clinical definition of each component of the syndrome using the BTPM, and in Section 4 we use the BDMM and compare the goodness of fit of both models.

Definition of the Metabolic Syndrome (MS)

The widely used definition of metabolic syndrome is that of the World Health Organization [16]. The components of each definition and criteria for making the diagnosis of the metabolic syndrome are summarized in Table 1. In addition, definitions were proposed by the European Group for the Study of Insulin Resistance [17] and the

American Association of Clinical Endocrinologists [18]. Finally, the (international Diabetes Federation (IDF) definition uses a lower fasting glucose level than the original NCEP definition, using the American Diabetes Association 2003 cut point for impaired fasting glucose [19,20].

Computation of MS prevalence from the prevalence of its individual components should be done using the basic principles of calculus of probability. This is because according to the clinicians’ definition, a randomly selected person is said to have the MS if he/she suffers from at least three of the five components. With this definition, we illustrate the calculations using the USA data given in Table 2.

As an example, the prevalence of one component is the probability assigned to the set:

$$S_1 = \{ \overline{A}BCDE \} U \{ A\overline{B}CDE \} U \{ ABC\overline{D}E \} U \{ ABCD\overline{E} \} U \{ ABCDE \}$$

Assuming the independence of the individual components, this is given by:

$$P(S_1) = P(\overline{A})P(B)P(C)P(D)P(E) + P(A)\overline{P(B)}P(C)P(D)P(E) + P(A)P(\overline{B})P(C)P(D)P(E) + P(A)P(B)\overline{P(C)}P(D)P(E) + P(A)P(B)P(\overline{D})P(E) + P(A)P(B)P(C)\overline{P(D)}P(E) + P(A)P(B)P(C)P(\overline{E})$$

The dash over the notation means the negation of the event. For example, $P(\overline{A}) = 1 - P(A) = 47\%$. Similar calculations give $P(S_i) = 0.283$. Following the above approach prevalence, of 2, 3, 4, and 5 components are given respectively as:

$$P(S_2) = \text{Prevalence of two components} = 0.0332, P(S_3) = \text{Prevalence of three components} = 0.213$$

$$P(S_4) = \text{Prevalence of four components} = 0.060, P(S_5) = \text{Prevalence of five components} = 0.006,$$

$$\text{and } P(S_0) = 1 - \sum_{i=1}^5 P(S_i) = 0.087.$$

Since in a large population MS prevalence is defined as the probability that a randomly selected individual has at least three components, this probability is estimated by:

$$P(\text{MS}) = P(S_3) + P(S_4) + P(S_5) = 0.006 + 0.060 + 0.213 \approx 28\%$$

On average, the expected number of MS components is $\sum_{i=0}^5 i P(S_i) = 1.25$. That is in 2006, for a randomly selected adult American, we would expect him/her to have on the average one component.

Component	IDF		WHO		EGIR		NCEP	
	M	F	M	F	M	F	M	F
A. Central Obesity	≥ 102	≥ 88	≥ 102	≥ 88	≥ 94	≥ 80	≥ 102	≥ 88
B. Raised TG	≥ 1.7	≥ 1.7	≥ 1.7	≥ 1.7	≥ 2.0	≥ 2.0	≥ 1.7	≥ 1.7
C. Low HDL	<1.03	<1.29	≤ .9	≤ 1	≤ 1	≤ 1	≤ 1.03	≤ 1.29
D. Hypertension	≥ 130/85	≥ 130/85	≥ 140/90	≥ 140/90	≥ 140/90	≥ 140/90	≥ 130/85	≥ 130/85
E. Fasting Glucose	≥ 5.6	≥ 5.6	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1

Table 1: Definition of the components of metabolic syndrome [Journal of the Royal Society of Medicine Vol. 99 September 2006]. IDF: International Diabetes Federation; WHO: World Health Organization; EGIR: European Group for the Study of Insulin Resistance; NCEP: National Cholesterol Education Program.

Component	A (central obesity)	B (TG)	C (low LDL)	D (hypertension)	E (fasting glucose)
Prevalence	53%	31%	25%	40%	39%

Table 2: Prevalence of Components of MS in USA-2006.

Spousal Concordance of Metabolic Syndrome: Analysis Based on Modeling Interval-Scale Data

An important study was conducted in [15]. The main objective was to estimate the spousal concordance for each of the 5 components of the syndrome. The sample included 3141 Korean spousal pairs. As a measure of concordance the authors used a single estimate of Pearson's correlation between spouses for each component. The data were not available for reanalysis by us. However, we used several Monte Carlo simulations from bivariate normal distributions. We used as input parameters the estimates reported in the Korean study with a sample size $n=4000$ spousal pairs. When we contrasted the summary measures in Table 3 with corresponding estimates produced by Park et al. [15] there were no differences among their and our results. In family studies, the ICC, not Pearson's correlation is the most commonly used measure of clustering of traits. In the last column of Table 3 we added the estimates of the Intra-class Correlation Coefficients (ICC) denoted by $\hat{\rho}$. Equation (1) shows how the ICC is obtained as a function of the means, variances and the Pearson's correlation.

In [15] the authors dichotomized the data for each component, using the WHO definitions, and then used the odds ratio as a measure of spousal concordance. Alternatively we shall combine the ICC estimates to produce an overall estimate of clustering of the components of the syndrome.

In equation (1) r denotes Pearson's correlation (\bar{x}_m, \bar{x}_f) are the mean scores for wife and husband, while (s_m, s_f) are the standard deviations of the scores. Note that that ρ is always less than r , and that $r=\rho$ only when $\bar{x}_m = \bar{x}_f$ and s_m, s_f .

Note that the ICC is quite low in all the components of the metabolic syndrome. Since the ICC is defined as the ratio of the between pairs variance to the total variance, its interpretation depends on the population under study. In societies where for example consanguinity is accepted, due to the larger contribution of the genetic components of variation, one would expect the value of the ICC to be much larger than what is reported here.

To find an overall estimate of a measure of clustering of the metabolic syndrome we follow an approach proposed by Cochran [21].

Let $w_i = v_i^{-1}$ denote the inverse of the variance of the MLE $\hat{\rho}_i$. Following Cochran [21], the minimum variance linear estimator of the pooled estimator of ICC is:

$$\bar{\rho} = \frac{\sum_{i=1}^5 w_i \hat{\rho}_i}{\sum_{i=1}^5 w_i} \text{ . We can show that the variance of } \bar{\rho} \text{ is given by:}$$

$$\text{var}(\bar{\rho}) \equiv v = \left(\sum_{i=1}^5 w_i \right)^{-1} \text{ .} \tag{2}$$

Asymptotically, $\bar{\rho}$ is distributed as $\bar{\rho} \sim N(\rho, v)$. Based on the asymptotic distribution of $\bar{\rho}$, we construct a test of homogeneity for the five ICC's. This test is based on X^2 and is formulated as follows:

$H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = \rho$ versus H_1 : at least two ρ 's are not the same. The null hypothesis is rejected if $T = \sum_{i=1}^5 w_i (\hat{\rho}_i - \bar{\rho})^2$ exceeds the chi-square tabulated value with 4 degrees of freedom and a given level for Type I error rate. For the simulated data, the estimated ICC and their standard errors are in Table 3. From results given in Donner [22] we have $\text{var}(\hat{\rho}_i) = (1 - \rho_i^2)^2 / k$, $i = 1, 2, 3, 4, 5$, we get $\bar{\rho} = 0.0967$, and $SE(\bar{\rho}) = .007$.

The value of the chi-square test $T=16.39$ giving a p -value=0.005. This means the level of clustering varies significantly across the components. We may construct a 95 % confidence interval in the form $.0967 \pm 1.96 (0.007)$.

This information can be used to estimate the sample size (number of spousal pairs) needed to verify relevant hypotheses on the clustering parameter.

Suppose, for example that we would like to estimate the number of spousal pairs needed to detect the departure of the common clustering parameter from its null value. This is equivalent to testing the null hypothesis: $H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$ v.s. H_1 : at least two ρ 's are not the same. We suggest applying the Fisher's [21] variance stabilizing transformation to the estimators of ICC. Fisher showed that, when the sample size is large, then for a sample size of N spousal pairs:

$$\varphi_i = \frac{1}{2} \log \left(\frac{1 + \hat{\rho}_i}{1 - \hat{\rho}_i} \right) \sim N \left(\frac{1}{2} \log \left(\frac{1 + \rho_i}{1 - \rho_i} \right), \frac{1}{k} \right), i=1, 2, 3, 4, 5 \tag{3}$$

With power 80% and Type I error rate 5%, to verify the above null hypothesis we need to recruit approximate number of pairs:

$$k = \frac{20(1.64 + 0.84)^2}{\left[\sum_{i=1}^5 \log \left(\frac{1 + \rho_i}{1 - \rho_i} \right) \right]^2} \tag{4}$$

For example for, $\rho_1 = .01, \rho_2 = .02, \rho_3 = .05, \rho_4 = .1$, and $\rho_5 = .2$, then $k=225$ pairs should be recruited. On the other hand, $\rho_1 = .01, \rho_2 = .01, \rho_3 = .01, \rho_4 = .01$, and $\rho_5 = .01$, then $k=13200$ pairs are needed. In practice, one may consider an average value for the ICC under the alternative, ρ^* . Substituting in equation (4) the estimated number of clusters would be:

$$k = \frac{20(1.64 + 0.84)^2}{\left[5 \log \frac{1 + \rho^*}{1 - \rho^*} \right]^2} \tag{5}$$

The first and second scenarios give sample sizes 212 and 12300.

Analysis Based on Clinical Definition of the Syndrome

The dichotomization scheme

The classification of a subject to be belonging to either of the two

MS Component	Father $\bar{x}_f \pm s_f$	Mother $\bar{x}_m \pm s_m$	r (Pearson's correlation)	ICC \pm SE
A. Central Obesity	84.3 \pm 8.2	78.8 \pm 8.9	0.06	0.05 \pm .016
B. TG	1.67 \pm 1.5	1.3 \pm .73	0.10	0.09 \pm .016
C. HDL	1.19 \pm .31	1.28 \pm .29	0.11	0.10 \pm .016
SBP	127 \pm 18.6	120.8 \pm 19.2	0.11	0.10 \pm .016
DBP	81.2 \pm 11.6	75.7 \pm 11.2	0.13	0.12 \pm .016
E. Fasting Glucose	5.6 \pm 1.5	5.5 \pm 1.6	0.08	0.07 \pm .016

Table 3: Mean \pm SD and intraclass correlation coefficient (ICC).

categories depends on the cut-off point specified by the adopted guide lines specified by the organizations (WHO, IDF, EGIR, NCEP). There are little differences among these thresholds and for our study, whose main objective is the statistical methods post categorizations, we shall use the recommendation made by the WHO. To formalize the presentation we suppose that y_{ijl} is the value of the l th component, for the j th spouse in the i th family, where, $i=1, 2, \dots, k$ (number of families), $j=1, 2$, and, $l=1, 2, 3, 4, 5$ We shall further assume that:

$$y_{ijl} = \mu_{jl} + \epsilon_{ijl}, \text{ and} \quad (6)$$

$$\epsilon_{ijl} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1l}^2 & r_l \sigma_{1l} \sigma_{2l} \\ r_l \sigma_{1l} \sigma_{2l} & \sigma_{2l}^2 \end{pmatrix}\right), \text{ for } l=1, 2, 3, 4, 5 \text{ and } j=1, 2,$$

Let

$$y_{ijl}^* = \begin{cases} 1 & \text{if } y_{ijl} > c_{jl} \\ 0 & \text{if } y_{ijl} < c_{jl} \end{cases}$$

Where c_{jl} are the threshold value of the l th component for the j th spouse, as determined by the IDF (International Diabetes Federation). Note that the inequality sign will be reversed for the case low HDL. The cumulative probit model for the categorical variable y_{ijl}^* can be written as:

$$P_{jl} = P_r(y_{ijl}^* = 1) = 1 - \Phi\left(\frac{c_{jl} - \mu_{jl}}{\sigma_{jl}}\right) \quad (7)$$

$j=1, 2; l=1, 2, 3, 4, 5$

$$\text{and } P_{j5} = \Phi\left(\frac{c_{j5} - \mu_{j5}}{\sigma_{j5}}\right) \text{ for low HDL.}$$

It turns out that, conditional on the i th spousal pair that $\{y_{ijl}^*\}$ constitutes a finite sequence of independent Bernoulli trials with success probability P_{jl} . Since we are interested in the sum of the components of MS, we define another random variable.

$$z_{ij} = \sum_{l=1}^5 y_{ijl}^*, \quad j = 1, 2; i = 1, 2, \dots, k$$

Note that $z_{ij} = 0, 1, 2, 3, 4$ and 5

Categorical data analysis

Fekedulegan et al. [23] conducted a comparison among several statistical models to evaluate factors associated with metabolic syndrome. The goal of their study was to evaluate the usefulness of alternative generalized linear models for analysis of metabolic syndrome as a count outcome and compare the results with models that utilize the definition as a binary outcome (presence/absence) for the syndrome. After the dichotomization of the measured outcomes, the definition of metabolic syndrome can be modified, as the total count of syndrome components for an individual, to represent a discrete outcome $y=0, 1, 2, 3, 4$, or 5 , where statistical models for count data can be used as an alternative to assess the association between exposure variable(s) and metabolic syndrome. They proposed using the Poisson regression to model the relationship between the syndrome as a dependent variable with limited range (0, 1, 2, 3, 4, 5) and potential risk factors associated with the syndrome. We note that:

1. The Poisson random variable has values that range from zero to infinity, and one has to modify the model to account for the right truncation at $y=5$.

2. Since our interest is on the clustering, a right truncated

Poisson model needs to be extended to model the correlation among the counts of the spousal pairs.

The focus of this section is on the application of the right truncated bivariate Poisson as a possible tool to model correlated count data.

Let x_l be a random variable having a truncated Poisson distribution, at a known truncation point m :

$$P(x_l = x) = \bar{e}^{-\lambda} \lambda^x / (x! \varphi_m(\lambda_l)) \quad l = 1, 2, \quad (8)$$

$$\text{where, } \varphi_m(\lambda_l) = \sum_{j=0}^m \bar{e}^{-\lambda} \lambda_l^j / j! \quad x_l = 0, 1, 2, \dots, m. \quad (9)$$

Denote the mean and variance of x are given respectively by μ_{m1} and v_{m1} . Simple computations show that the mean and variance are given respectively by:

$$E(x_l) = \mu_{m1} = \left[1 + \frac{\varphi'_m(\lambda_l)}{\varphi_m(\lambda_l)}\right] \quad (10)$$

$$\text{var}(x_l) = v_{m1} = \lambda_l^2 \left[\frac{\varphi''_m(\lambda_l)}{\varphi_m(\lambda_l)} - \left(\frac{\varphi'_m(\lambda_l)}{\varphi_m(\lambda_l)} \right)^2 \right] + \lambda_l \left[1 + \frac{\varphi'_m(\lambda_l)}{\varphi_m(\lambda_l)} \right] \quad (11)$$

Strategy of the construction of a bivariate model from the marginal distributions is a well-known [24, 25].

Define the random variables:

$$\xi_l = x_l + \gamma,$$

where (x_1, x_2, γ) are mutually independently Poisson random variables with $\gamma \sim \text{Poisson}(\theta)$.

The random variables (ξ_1, ξ_2) are said to have bivariate truncated Poisson distribution whose joint probability function is given by:

$$P(\xi_1 = r, \xi_2 = s) = K \lambda_1^r \lambda_2^s \sum_{\gamma=0}^{\min(r,s)} \frac{(\theta / \lambda_1 \lambda_2)^\gamma}{\gamma! (r-\gamma)! (s-\gamma)!} \quad (12)$$

$r, s = 0, 1, 2, 3, 4, 5$, and $\gamma = \min(r, s)$, and

$$K = \exp(-\lambda_1 - \lambda_2 - \theta) \left[\sum_{r=0}^5 \sum_{s=0}^5 P(\xi_1 = r, \xi_2 = s) \right]^{-1} \quad (13)$$

It can be easily shown that ϑ is the covariance between the two variables and therefore:

$$\text{corr}(\xi_1, \xi_2) = \frac{\vartheta}{\sqrt{(\vartheta + v_{m1})(\vartheta + v_{m2})}} \quad (14)$$

Because of the restrictions on the Poisson parameters, the correlation between (ξ_1, ξ_2) is strictly positive. We followed the WHO guidelines as indicated in Table 1, to dichotomize the data obtained from the Monte Carlo simulations.

Let the estimated count frequency from the 5 X 5 table be denoted by $n_{rs} = n \hat{P}(\xi_1 = r, \xi_2 = s)$. Here $\hat{P}(\cdot, \cdot)$ are the estimated probabilities after substituting the moment estimators $\hat{\lambda}_1 = 1.336, \hat{\lambda}_2 = 0.986, \text{ and } \hat{\theta} = .061 \pm 0.012$. Maximum likelihood estimators are possible to obtain, by solving numerically the likelihood equations where the likelihood function is given by:

$$L = \prod_{r=0}^m \prod_{s=0}^m [P(\xi_1 = r, \xi_2 = s)]^{n_{rs}} \quad (15)$$

The correlation coefficient estimate is obtained by substituting the MLE of the model parameters in (4) and is given by $\hat{\rho} = 0.0798 \pm 0.016$. The standard error of $\hat{\rho}$ is obtained by the delta method. The concordance correlation, as a measure of clustering, using equation (1) is $\hat{\rho}_c = 0.067 \pm 0.0161, p\text{-value} = 6.8 \times 10^{-6}$.

Note that the cost of categorization, relative to the uncategorized data is a drop of 29% in the value of the clustering parameter (0.0760 vs 0.0967) and a drop in the efficiency of estimation of about 19%. Meaning that, if we follow the clinical definition of the syndrome, one needs a sample size almost 20% larger than the needed sample to analyze uncategorized data, to maintain the level of precision in estimating the clustering parameter.

An important question is how to assess the goodness-of-fit (GOF) of the BTPM. Traditionally the chi-square goodness of fit is used for this purpose. But one of the disadvantages of this measure is that its value is affected by few (one or two) expected frequencies (e) that deviate markedly from their corresponding observed frequencies (o). Alternatively, we shall use three approaches to measure the GOF of the BTPM. The first is to use Lin's concordance correlation between the observed, and the expected counts n_{rs} under the present model. Using equation (1) we get an estimate measure of concordance between the observed and the expected counts given by (0.798 ± 0.091). This is an indication that the BTPM fits the counts reasonably well. The second is quite different, and depends on the degree of closeness (agreement) between the observed and the expected counts under the present model. First we define the statistic $D = \text{abs}(o-e)/25$, and if $D \leq q$ then, e and o are close, otherwise they are not. The quantity q is arbitrarily chosen, and for the current application we take $q=0.01$. The percent agreement (i.e. percent of time $D \leq 0.01$) is 0.375, with a standard error 0.125, indicating poor agreement between the observed and the expected counts. The third approach is an adaptation to the technique developed by Altman and Bland [26] and Bradley and Blackwood [27] to establish agreement between two sets of measurements. As shown in [27] a non-significant liner regression of (o-e) on (o+e)/2 is an indicative of strong agreement. In the meant time, we can detect agreement as indicated in [25] by plotting the limits of agreement:

$(o-e) \pm SD(o-e) \text{ vs } (o+e)/2$. This is shown in Figure 1, and we

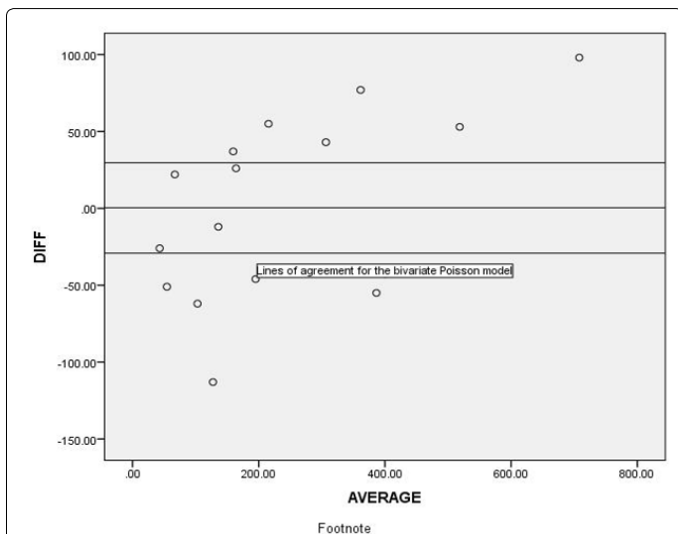


Figure 1: Limits of agreement between the observed and expected counts based on the BTPM.

conclude, based on the methods that the BTPM does not provide a good fit to the data.

The BTPM allows us to analyze the categories as counts, which are often modeled by probability distributions such as Poisson or negative binomial. Another strategy is to analyze the frequency of these counts in the sampled spousal pairs. The most commonly used tool that accounts for the correlations among categories for cross classified data is the Multinomial-Dirichlet-Model (MDM) [28,29]. As a competitor to the BTPM we shall apply the MDM to analyze correlated categorical data.

Analysis of Correlated Categorical Data with the BDMM

To analyze the cell counts in the 6×6 frequency table assuming that the data were generated by the MDM, we need to formalize the presentation.

Conditional on the i^{th} pair (z_{i1}, z_{i2}) have independent multinomial distribution with:

$$P(z_{ij} = x) = \eta_{ij}(x) \tag{16}$$

$$\eta_{ij}(0) = P_r[z_{ij} = 0] \tag{17}$$

$$= \prod_{l=1}^5 \left[1 - \Phi \left(\frac{c_{jl} - \mu_{jl}}{\sigma_{jl}} \right) \right]$$

Now, to affect a correlation among the multinomial probabilities, we usually assume that conditional on the spousal pair the vector of probabilities $(\eta_{i1}^{(a)}, \eta_{i2}^{(b)})$ for $a=0,1,2,3,4,5$ and $b=0,1,2,3,4,5$ has a Dirichlet distribution [25] as a mixing distribution. Instead, we shall assume that $(\eta_{i1}^{(a)}, \eta_{i2}^{(b)})$ has a bivariate distribution that is specified by its marginal means and the covariance structure such that:

$$1. E(\eta_{ij}^{(a)}) = \pi_j^{(a)}, \text{ and } E(\eta_{ij}^{(b)}) = \pi_j^{(b)} \tag{18}$$

$$2. \text{var}(\eta_{ij}^{(a)}) = \rho \pi_j^{(a)} (1 - \pi_j^{(a)}) \quad j = 1, 2, \text{ and} \tag{19}$$

$$3. E(\eta_{i1}^{(a)} \eta_{i2}^{(b)}) = \pi_i^{(a)} \pi_j^{(b)} + \rho \sqrt{\pi_i^{(a)} \pi_i^{(a)} \pi_j^{(b)} \pi_j^{(b)}} \quad i \neq j = 1, 2 \tag{20}$$

where, and ρ is the correlation between $\pi_i^{(a)}(\cdot) = 1 - \pi_i^{(a)}(\cdot)$ and will be treated as a nuisance parameter. Note that the correlation parameter defined is not necessarily the clustering parameter that we seek to estimate. Introducing ρ at this stage is necessary to create a quasi-random effect model in order to induce correlation between spousal pairs. The idea is that we need to generate expected correlated cell counts under the above set-up and then compare these counts with the observed counts. As a measure of cluster we shall use Cohen's kappa [30,31].

The application of Cohen's Kappa in clinical and epidemiological research is long standing and is used to measure the diagonal agreement between two categorical variables in $C \times C$ table. Let

$$\pi_0 = \sum_{i=1}^c \pi_{ii} \text{ and } \pi_e = \sum_{i=1}^c \pi_i \cdot \pi_i \text{ where } \pi_0 \text{ represents the observed proportion of diagonal counts, and } \pi_e \text{ represents the expected proportion of diagonal counts if the } z_{i1} \text{ and } z_{i2} \text{ are independent.}$$

Note that

$$\pi_{i.} = \sum_{j=0}^5 \pi_{ij}, \pi_{.j} = \sum_{i=0}^5 \pi_{ij}, \text{ and } \sum_{i=0}^5 \pi_{i.} = \sum_{j=0}^5 \pi_{.j} = \sum_{i=0}^5 \sum_{j=0}^5 \pi_{ij} = 1$$

The kappa coefficient is defined as:

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \quad (21)$$

Let $p_0 = \sum_{i=1}^c p_{ii}$ and $p_e = \sum_{i=1}^c p_i \cdot p_i$. Here, $p_{ij} = \frac{n_{ij}}{N}$ are point moment estimators of π_{ij} , and p_0 represents the observed proportion of diagonal counts, and p_e represents the expected proportional of diagonal counts if the x and y are independent. The estimated Kappa is:

$$\tilde{\kappa} = \frac{p_0 - p_e}{1 - p_e} \quad (22)$$

Agresti [29] presented a simple expression for the estimated variance of $\tilde{\kappa}$

$$\widehat{\text{var}}(\tilde{\kappa}) = \frac{1}{N} \left[\frac{p_0(1-p_0)}{(1-p_e)^2} + \frac{2(1-p_0)c}{(1-p_e)^3} + \frac{(1-p_0)^2 d}{(1-p_e)^4} \right] \quad (23)$$

Where;

$$c = 2p_0 p_e - \sum_{i=1}^c p_{ii} (p_i + p_i) = 2p_0 p_e - u_1$$

$$d = \sum_{i=1}^c \sum_{j=1}^c p_{ij} (p_i + p_j)^2 - 4p_e^2 = u_2 - 4p_e^2$$

It is more convenient to write the variance of $\tilde{\kappa}$ in terms of the chance agreement and the population kappa in the form:

$$\widehat{\text{var}}(\tilde{\kappa}) = \frac{1 - \tilde{\kappa}}{N(1 - p_e)} \left[\tilde{\kappa} + p_e(1 - \tilde{\kappa}) + \left(\frac{4p_e \tilde{\kappa} - 2u_1}{1 - p_e} + (1 - \tilde{\kappa})u_2 \right) \right] = \psi(\kappa) / N \quad (24)$$

We would like to determine the sample size N needed to test the hypothesis

$$H_0: \kappa=0 \text{ verses } H_1: \kappa=1$$

Type I error date= α and Type II error date= β the required number of spousal pairs is

$$N = \left[\frac{Z_\alpha \sqrt{\psi(\kappa_0)} + Z_\beta \sqrt{\psi(\kappa_1)}}{\kappa_0 - \kappa_1} \right]^2 \quad (25)$$

Remarks:

When $\kappa=0$, then

$$\begin{aligned} \text{Var}(\hat{\kappa} | \kappa = 0) &= \frac{1}{N} \left(\frac{\pi_e}{1 - \pi_e} + \frac{u_2 - 2u_1}{(1 - \pi_e)^2} \right) \\ &= \frac{1}{N} \left[\frac{1}{1 - \pi_e} - 1 + \frac{u_2 - 2u_1}{(1 - \pi_e)^2} \right] \end{aligned} \quad (26)$$

Therefore $\text{Var}(\hat{\kappa} | \kappa = 0)$ is an increasing function of π_e . That is, larger values of chance agreement requires larger number of spousal pairs to detect significant departure from the null value. We also note that the complete specifications of $\text{Var}(\hat{\kappa})$ depend on values of u_1 and u_2 which in turn depend on the cell and the marginal probabilities. Therefore one should provide guessed values for u_1 and u_2 in an attempt to estimate the sample size.

We shall use Cohen's [30,31] kappa to measure the degree of clustering between randomly selected pairs of spouses. Borkowf et al. [32] established the asymptotic theory for construction of large sample confidence interval on the kappa statistic.

The cross classification of the data for wives and husbands is given in Table 4.

Note that category 5 was quite sparse for both males and females and therefore we decided to collapse it with category 4. The estimated κ and its standard errors are given respectively by 0.022, and 0.01, which is significantly different from zero (p-value=0.027). The BDMM fitted the data quite well. The first criterion for the goodness of fit gives: (1) an estimated value for Lin's concordance correlation coefficient between observed and expected count as (0.995 ± 0.002) . (2) The percentage absolute difference (percentage agreement), using the same criterion as in the case of the BTM is (0.92 ± 0.06) . (3) The limits of agreement [26] shown in Figure 2 indicate that there is a considerable agreement between the observed and expected counts. The regression of the difference between the observed counts on the average was non-significant with p-value=0.133.

Marginal homogeneity

In order to establish familial resemblance between pairs of spouses it is important to test the homogeneity of their marginal distributions. In the case of the matched-pairs data, McNemar's test [33] can be applied only to the case in which there are two possible categories for the outcome. According the clinical definition of the MS a person with at least three components is classified to have the condition; else he/she does not have it. For spousal pairs with a categorical response, a two-way contingency table with the same row and column categories summarizes the data, under this situation, the contingency table is also called square table. In this case we have single 2×2 table:

The variance expression for kappa (21) is greatly simplified, and

		Female					Total
		0	1	2	3	4	
Male	0	141 (139)	285 (295)	218 (235)	48 (77)	8 (13)	700
	1	323 (310)	659 (673)	471 (533)	167 (180)	17 (26)	1637
	2	188 (235)	492 (509)	414 (404)	113 (130)	21 (21)	1228
	3	56 (82)	151 (174)	142 (140)	49 (47)	6 (8)	404
	4	7 (9)	8 (16)	11 (13)	5 (5)	0 (1)	31
Total		715	1595	1256	382	52	4000

Table 4: Cross classification counts of the 5 components. The bracketed numbers are the expected frequencies $e=n_{is}$ based on the MDM where the bolded numbers are the observed counts.

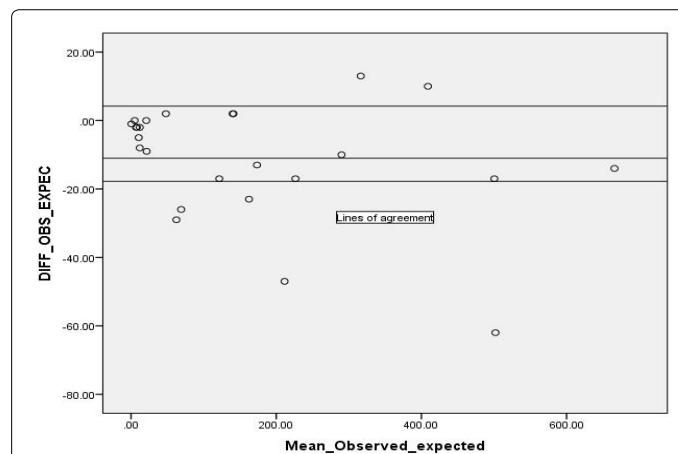


Figure 2: Limits of agreement between the observed and expected counts based on the BDMM.

will depend on the marginal probabilities and kappa:

$$\pi_{1.} = 1 - \pi_{0.}, \pi_{.1} = 1 - \pi_{.0}, \pi_e = \pi_{0.}\pi_{.0} + \pi_{1.}\pi_{.1}, \pi_0 = \kappa(1 - \pi_e) + \pi_e$$

$$\pi_{11} = (\pi_{0.} - \pi_{0.}\pi_{.1}) / 2, \pi_{00} = \pi_{0.} - \pi_{11}, \pi_{01} = \pi_{0.} - \pi_{00}, \text{ and } \pi_{10} = \pi_{.0} - \pi_{00}$$

Under marginal homogeneity the variance expression for the kappa statistic reduces to:

$$Var(\hat{\kappa}) = \frac{1 - \kappa}{N} [(1 - \kappa)(1 - 2\kappa) + \frac{\kappa(2 - \kappa)}{2\pi(1 - \pi)}] \quad (27)$$

Following the clinical definition of the MS, we collapse the data in Table 5 into a square Table 6. The response category “1” represents the count of at least three components.

The McNemar’s test of marginal homogeneity has a p-value=1.00. The kappa statistic and its standard error are given respectively by (0.033 ± 0.017). There is some statistical implication when the clinical definition of the syndrome is used.

For a general square contingency table Stuart [34] and Maxwell [35] developed an asymptotic chi-square test of marginal homogeneity, for which a SAS macro was provided in [36]. The value of the test statistic is 16.3 with p-value=0.092 indicating the support for marginal homogeneity between spouses’ distribution over the specified categories.

The issue of sample size requirements

The sample estimation is an essential step to guarantee that a study possesses a certain power. In the case of 2 × 2 table, sample size requirements regarding the statistical inference on kappa, has been extensively discussed [37,38]. In the general case, we shall investigate two approaches. The first approach is proposed by Cantor [39] who used the relative error in estimation ϵ and the numerator of kappa ($\pi_{0.} - \pi_e$) as the basic requirements for sample size estimation. The proposed equation is:

$$N = \frac{1}{\epsilon^2 (\pi_{0.} - \pi_e)^2}$$

We modify the above equation such that:

$$N = \frac{1}{\epsilon^2 \kappa^2 (1 - \pi_e)^2} \quad (28)$$

Calculations based on the above equation gives very unrealistic sample size estimates. For example when, $\pi_e = 0.30, \pi_{0.} = 0.5$, implying that kappa=0.29, and making an error in estimation, $\epsilon = 0.05$, we need to recruit at least 9700 spousal pairs. Two advantages of Cantor’s

Husband	Wife		Total
	0	1	
0	π_{00}	π_{01}	$\pi_{0.}$
1	π_{10}	π_{11}	$\pi_{.1}$
Total	$\pi_{.0}$	$\pi_{.1}$	1

Table 5: The case of square 2 × 2 table.

husband		wife		Total
		0	1	
husband	0	3191	374	3565
	1	375	60	435
Total		3566	434	4000

Table 6: The collapsed data.

π_e	u_1	u_2	κ	
			.05	.10
0.01	0.01	0.01	133	40
0.05	0.1	0.1	966	68
0.1	0.05	0.05	753	193
0.01	0.1	0.1	800	204
0.4	0.01	0.1	2447	608
0.4	0.05	0.05	2654	660

Table 7: Sample size requirements to test the hypothesis $H_0: \kappa=0$ under the MDM.

expression; the first is that the sample size equation (28) does not depend on the marginal probabilities, through its dependence on u_1 and u_2 which must be guessed by the researcher. Second, it does not need specifications for the values of kappa under the two hypotheses to be tested. Alternatively, we provide the sample size requirement to test the hypothesis $H_0: \kappa=0$ versus a non-zero alternative, for Type I error rate α and power $1 - \beta$. The sample size expression is:

$$N = \left\{ \frac{z_{1-\beta} \sqrt{\psi(\kappa_1)} + z_{\alpha} \sqrt{\psi(\kappa_0)}}{\kappa_0 - \kappa_1} \right\}^2 \quad (29)$$

For testing $\kappa_0=0$, Table 7 gives some calculated values of N for guessed values of u_1 and u_2 , 80% power and 5% Type I error rate. The disadvantage is that for (29) to be of practical importance we need to guess the values of u_1 and u_2 .

It is clear from Table 7 that for fixed power and type error rate the sample size depends on the distance of the values of under both hypotheses, and the chance error. Moreover, it seems that values of u_1 and u_2 have little effect on the sample size.

Discussion

Our main focus in this paper has been on the estimation of spousal concordance which may be the result of shared lifestyle and socioeconomic environment. In fact spousal resemblance or concordance may be due to shared environment, common behaviors, and also positive assortative mating, that is, the tendency of individuals to choose a spouse with similar characteristics. If concordance was mainly due to a cohabitation effect, then it should increase with increasing time shared by spouses. Identification of the relative contributions of shared modifiable environmental risk factors may then improve our understanding and thus enable targeting of detrimental lifestyle minimizing the rapid increase in the prevalence of the metabolic syndrome.

Because data for the components of MS for spousal pairs are not available, we simulated similar data using input parameters from the Korean study. Working with simulated data is a practical way to control the sampling error, and to verify the reproducibility of the results. This strategy has been recommended in [40]. Since our interest is in modeling MS in spousal pairs we need to define bivariate probability distributions for continuous data, and the categorical data. We developed two models; (1) The Truncated Bivariate Poisson Model (BTPM) to analyze the categories as counts. The main purpose of the truncation is to overcome the limitations of the Poisson to model discrete outcomes that has an upper finite limit. (2) The second model is Multivariate-Dirichlet-Multinomial distribution (BDMM) to analyze the frequency of the bivariate counts. The two models are not nested within each other, and we therefore had to evaluate the goodness of fit of each model separately. We have developed three approaches to goodness of fit; all showed that the BDMM fitted the data quite closely.

Since the index of clustering is our target parameter its nature varied according to the modeling strategy. The ICCC is a natural choice for the continuous data under both the bivariate normal model and the BTPM. The kappa statistic is the natural choice to measure spousal concordance for the categorical data. It is interesting to notice that the parameter estimate of spousal concordance has almost similar value under the three models. But the level of uncertainty was highest under the BTPM.

There are limitations in modeling MS under the study design proposed in [15]. The first is the absence of important covariates, and or possible confounders, that affect the estimation of the clustering parameter. For example, it would be desirable to include in the proposed models individual level covariates such as age, and level of education, and possibly cluster level covariate such as the length of spousal co-habitation.

As a final remark, we note that the evaluation of the kappa statistic from a 2X2 table makes the problem of likelihood based inference, sample size estimation, and confidence interval construction tractable. But the interest in agreements between pairs when the subjects are classified in several categories has increased the potential applications of the kappa statistic in this regard as a measure of clustering. Hypothesis testing and sample size requirements are much more complicated in the multi-categorical classification. This problem was considered by many researchers by collapsing the multinomial data into binary data.

Bartfay and Donner [41], Cohen [42] Donner and Eliasziw [43], and Kraemer [44] demonstrated the advantage of preserving multinomial data on the original scale. They all showed that collapsing multinomial data into two categories results in reduction of the effective sample size. This means that a substantial increase in the sample size is required to maintain the same level of power at a given level of significance.

References

1. Laaksonen D, Lakka H, Niskanen L, Kaplan G, Salonen J, et al. (2002) Metabolic syndrome and the development of diabetes mellitus: application and validation of recently suggested definitions of the metabolic syndrome in a prospective cohort study. *Am J Epidemiol* 156: 1070-1077.
2. Lakka HM, Laaksonen DE, Lakka TA, et al. (2002) The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA* 288: 2709-2716.
3. Carmelli D, Cardon LR, Fabsitz R (1994) Clustering of hypertension, diabetes, and obesity in adult male twins: same genes or same environments? *American Journal of Human genetics* 55: 566-573.
4. Edwards KL, Newman B, Mayer K (1997) Heritability of factors of the insulin resistance syndrome in women twins. *Genetic Epidemiology* 14: 241-253.
5. Hong Y, Pederson NL, Brismar (1997) Genetic and environmental architecture of the features of insulin resistance syndrome. *American Journal of Human genetics* 60: 143-152.
6. Lee KE, Klein BE, Klein R (2003) Familial aggregation of components of metabolic syndrome in the Framingham Heart and Offspring Cohorts: Genetic Analysis Workshop Problem 1. *BMC Genetics* 4: S94.
7. Park HS, Park JY, Cho S-IL (2006) Familial aggregation of metabolic syndrome in Korean families with adolescents. *Atherosclerosis* 186: 215-221.
8. Borch-Johnson, K (2007) The metabolic syndrome in a global perspective. *Danish Medical Bulletin* 54: 157-159.
9. Reaven GM, Olefsky J, Farquhar JW (1972) Does hyperglycaemia or hyperinsulinaemia characterise the patient with chemical diabetes? *Lancet* i: 1247-1249.
10. Reaven G (1988) Role of insulin resistance in human disease. *Diabetes* 37: 1595-1607.
11. Laird, N, Lange C (2010) The fundamentals of Modern Statistical genetics. Springer, New York.
12. Morton NE (1974) Analysis of family resemblance. I. Introduction. *American Journal of Human genetics* 26: 318-330.
13. Rao DC, Volger PG, McGue M, Russell JM (1987) Maximum likelihood estimation of familial correlations from multivariate quantitative data on pedigrees: A general method and examples. *American Journal of Human Genetics* 41: 1104-1116.
14. Morton NE, MacLean CJ (1974) Analysis of family resemblance. III Complex segregation analysis of quantitative traits. *American Journal of Human genetics* 26: 489-503.
15. Kim HC, Kang DR, Choi KS, Nam CM, Thomas GN (2006) Spousal concordance of metabolic syndrome in 3141 Korean couples: a national survey. *Annals of Epidemiology* 16: 292-298.
16. Cameron AJ, Shaw JE, Zimmet PZ (2004) The metabolic syndrome: prevalence in worldwide populations. *Endocrinol Metab Clin N Am* 33: 351-375.
17. World Health Organization (1999) Definition, Diagnosis, and Classification of Diabetes Mellitus and Its Complications. Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva.
18. Balkau B, Charles MA (1999) Comment on the provisional report from the WHO consultation. *European Group for the Study of Insulin Resistance (EGIR). Diabet Med* 16: 442-443.
19. Einhorn D, Reaven GM, Cobin RH, Ford E, Ganda OP, et al. (2003) American College of Endocrinology position statement on the insulin resistance syndrome. *Endocrine Practice* 9: 237-252.
20. Genuth S, Alberti KG, Bennett P (2003) Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 26: 3160-3167.
21. Cochran WG (1943) Analysis of variance of percentages based on unequal numbers. *JASA* 38: 287-301.
22. Donner A (1986) A reviews of inference procedures for the intra-class correlation coefficient in the one-way random effects model. *International Statistical Review* 54: 67-82.
23. Fekeddulegan D, Andrew M, Violanti J, Hartley T, Charles L, et al. (2009) Comparison of statistical approaches to evaluate factors associated with metabolic syndrome. *The Journal of Clinical Hypertension* 12: 365-373.
24. Hamdan MA, Al-Bayati HA (1969) A note on the bivariate Poisson distribution. *American Statistician* 23: 32-33.
25. Johnson NL, Kotz S, Balakrishnan N (1997) *Discrete Multivariate Distributions*. J. Wiley, New York.
26. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two Methods. *Lancet* 307-310.
27. Bradley EL, Blackwood LG (1989) Comparing paired data: A simultaneous test for means and variances. *The American Statistician* 43: 234-235.
28. Chen JJ, Kodell RL, Howe RB, Gaylor DW (1991) Analysis of trinomial responses from reproductive and developmental toxicity experiments. *Biometrics* 46: 1049-1058.
29. Agresti A (1990) *Categorical data Analysis*. (3rd edition) Wiley & Sons Inc, New York.
30. Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 37-46.
31. Cohen J (1968) Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70: 213-220.
32. Borkowf CB, Gail MH, Carroll RJ, Gill RD (1997) Analyzing bivariate continuous data grouped into categories defined by empirical quantiles of marginal distributions. *Biometrics* 53: 1054-1069.
33. McNemar QA (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12: 153-157.
34. Stuart A (1995) A test for homogeneity of the marginal distributions in a Two-Way Classification. *Biometrika* 42: 412-416.
35. Maxwell AE (1970) Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry* 116: 651-655.

36. Sun XU, Yang Z (2008) Generalized McNemar's test for homogeneity of the marginal distributions. *SAS Global Forum, Department of Epidemiology and Biostatistics, USA.*
37. Donner A, Eliasziw M (1992) Goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation. *Statistics in Medicine* 11: 1511-1519.
38. Shoukri M (2010) Measures of inter observer agreement and reliability. (2nd edition), CRC Press/Taylor and Francis, Boca Raton Fld, USA.
39. Cantor AB (1996) Sample size calculations for Cohen's Kappa. *Psychological Methods* 1: 150-153.
40. Khondoker M, Dobson R, Skirro C, Simmons A, Stahl D (2013) A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Statistical Methods in Medical Research*.
41. Bartfay E, Donner A (2000) The effect of collapsing multinomial data when assessing agreement. *International Journal of Epidemiology* 29: 1070-1075.
42. Cohen J (1983) The cost of dichotomization. *Applied Psychological Measurements* 7: 249-253.
43. Donner A, Eliasziw M (1994) Statistical implications of the choice between a dichotomous or continuous trait in studies of inter observer agreement. *Biometrics* 50: 550-555.
44. Kraemer HC (1979) Ramification of a population model for κ as a coefficient of reliability. *Psychometrika* 44: 461-472.