An Assessment of Quadratic Inference Functions Method

Karesel Çıkarsama Fonksiyonları Yönteminin Değerlendirilmesi

Sara SERAHATI,[a]
Akbar BIGLARIAN,[b]
Enayatollah BAKHSHI[b]

[a]Obesity Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences,
[b]Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, IRAN

Geliş Tarihi/*Received:* 14.05.2013
Kabul Tarihi/*Accepted:* 05.07.2013

Yazışma Adresi/*Correspondence:*
Enayatollah BAKHSHI
University of Social Welfare and Rehabilitation Sciences, Department of Biostatistics, Tehran, IRAN
bakhshi@razi.tums.ac.ir

ABSTRACT Objective: The generalized estimating equations (GEE) which is a method for estimating and belongs to the population average models is used in correlated data analysis. For the GEE method researcher must make guess for the working correlation structure. A related method to GEE is the quadratic inference functions (QIF). We aimed to compare two methods via an illustrative example and to estimate the relationship between sex, marital status, age, educational level, smoking status and obesity in participants who aged 20-65 years old. **Material and Methods:** This study compares two methods to estimate the odds of obesity (BMI ≥30 kg/m^2) as a function of age, sex, marital status, educational level and smoking status by using data from" Tehran Lipid and Glucose Study (TLGS)" database which include 1106 households and 3203 participants aged 20-65 years. **Results:** The odds ratio estimates for two methods changed only slightly but the relative efficiency of parameter estimates from GEE and QIF was 1.23. QIF can produce sample size saving for a given power. Two methods showed that increased age, being a nonsmoker, lower educational level, and being married, as well as female sex were positively associated with obesity. **Conclusion:** According to the results of this study, QIF is better than GEE in the case of misspecification of the working correlation structure and provides more efficient parameter estimates than GEE method.

Key Words: GEE; QIF; odds ratio; BMI; obesity

ÖZET Amaç: Bir tahmin yöntemi olan ve anakütle ortalama modellerine dayanan genelleştirilmiş tahmin denklemleri (Generalized Estimating Equations-GEE) ilişkili veri analizinde kullanılır. GEE yöntemi için araştırmacı, çalışan korelasyon yapısı için tahmin yapmalıdır. GEE ile ilgili olan bir yöntem karesel çıkarsama fonksiyonlarıdır (Quadratic Inference Functions-QIF). Bu iki yöntemi açıklayıcı bir örnek ile karşılaştırmayı ve yaşları 20-65 yıl arasında olan katılımcılarda cinsiyet, medeni durum, yaş, eğitim düzeyi, sigara içme durumu ve obezite arasındaki ilişkiyi tahmin etmeyi amaçladık. **Gereç ve Yöntemler:** Bu çalışma, yaşları 20-65 yıl arasında olan 1106 hane ve 3203 katılımcı içeren "Tahran Lipid ve Glukoz Çalışması" veritabanından gelen verileri kullanarak, yaş, cinsiyet, medeni durum, eğitim düzeyi ve sigara içme durumunun bir fonksiyonu olarak obezitenin (BMI ≥30 kg/m^2) oddsunun tahmin edilmesinde iki yöntemi karşılaştırmaktadır. **Bulgular:** İki yöntem için odds oranı tahminleri sadece çok az bir miktarda değişmiştir; fakat GEE ve QIF'nin parametre tahminlerinin göreli etkinliği 1.23'dür. QIF, verilen bir güçte örneklem büyüklüğü tasarrufu sağlayabilmektedir. İki yöntem, artan yaşın, sigara içicisi olmamanın, düşük eğitim düzeyinin, evli olmanın ve yanı sıra kadın cinsiyete sahip olmanın obezite ile pozitif yönde ilişkili olduğunu göstermişlerdir. **Sonuç:** Bu çalışmanın bulgularına göre; çalışan korelasyon yapısının yanlış belirlenmiş olması durumunda QIF, GEE'den daha iyidir ve GEE yönteminden daha etkin parametre tahminleri vermektedir.

Anahtar Kelimeler: GEE; QIF; odds oranı; BMI; obezite

Turkiye Klinikleri J Biostat 2014;6(1):1-7

In analyzing data the choice of an appropriate statistical technique is often straightforward. But, there are some circumstances in which this decision makes problems for many researchers. This problem can be particularly difficult for researchers whose work focuses on analyzing data that arise from correlated observations. For example, researchers encounter occasions in which outcomes may be observed in subjects over several time intervals (where individuals repeatedly measured over time) or they may be observed in clusters where multiple measurements are collected from the same experimental unit (e.g., Husband-wife pairs, siblings). These occasions cannot be considered as 'independent'. So, this lack of independence means that many traditional methods such as ordinary least squares (OLS) cannot be used to analyze data since one key assumption of OLS regression is that all observations are independent of each other.[1] Moreover, outcomes which are not distributed normally such as count data (e.g., Number of symptoms) require a more specialized analytical tool as well.[2]

Practical methods with reasonable statistical efficiency to analyze such data are marginal models, mixed effects models and transitional models.[3-5]

Among them, the purpose of marginal models is to estimate the population-average effect of covariates on the response of interest. The term marginal means that the model for mean response depends only on the covariates of interest, not on any random effects or previous responses.[6]

The generalized estimating equations (GEE) method is the most popular approach in marginal models which can handle a variety of correlated models. Additionally, GEE can handle types of outcomes (e.g., Count, binary, continuous) just as like time-varying and time-invariant predictors.[2] Over the past 20 years, GEE has been an exceedingly useful approach for analyzing longitudinal data[2]. The parameter estimates obtained from GEE method are consistent even if the true working correlation structure is misspecified. But, if this structure is correctly specified they are efficient in the sense that the asymptotic variance of parameter estimators attain a Cramer-Rao-type lower band.[7]

Wang and Carey have shown that appropriate specification of correlation structure in longitudinal data analysis improves efficiency of parameter estimates and leads to more reliable statistical inferences.[8]

Prentice and Zhao (1991) proposed an ad hoc method (GEE1) which introduces additional estimating equations for the working correlation Parameter.[9] Fitzmaurice et al.[7] showed that in order to improve the efficiency of the regression coefficients in quasi-likelihood inference, it is necessary to specify the working correlation matrix which is as close as possible to the true one. Odueyungbo et al. compared GEE and quadratic inference functions (QIF) using data from the National Longitudinal Survey of Children and Youth (NLSCY) with assuming AR-1 and exchangeable working correlation structures and showed that the estimators from QIF are more efficient than GEE.[5]

The GEE approach has other shortcomings such as lack of goodness of fit test, sensitivity to outliers or contaminated data and lack of model selection criteria like AIC and BIC. So, for overcoming the difficulties in using GEE, many improvements has been proposed. But among them as an alternative approach to GEE, Qu et al. proposed QIF method. This method is mostly used to improve the efficiency of parameter estimates when the working correlation structure is misspecified.[4,10]

The primary aim of this study was to compare the results from GEE and QIF by analyzing the data from families who participated in the Tehran Lipid and Glucose Study (TLGS),[11] and secondly to estimate the relationship between sex, marital status, age, educational level, smoking status and obesity in participants who aged 20-65 years old.

MATERIAL AND METHODS

GEE METHOD

GEE can be regarded as a multivariate extension of the quasi-likelihood estimating equations. This approach describes a marginal mean model in which the mean response among clusters (subjects, or between responses at different time periods) changes with covariates. Additionally, when response correlations are treated as a nuisance parameter, GEE method demonstrates the relationship between covariates and the probability of response.[7,12,13] This correlation apparently is modeled by assumption of a correlation structure (or a working correlation matrix) such as:

1) Independence working correlation: this model is the multivariate extension of ordinary logistic regression which assumes that repeated responses are independent.

2) Exchangeable: in this structure just one correlation parameter ρ is estimated which gives correlation of each pair of repeated measures (responses).

3) Unstructured: correlations within any 2 responses are unknown and need to be estimated. There are few limitations in this structure. But, if there has been much parameters it cannot become reliable.

4) Auto-regressive of order 1 (AR (1)): this case is more frequent among the others. In this case, just one correlation parameter has estimated and the correlation parameter has been distinct by the lag of time between responses which is assumed to be the same between any two observations and measures further apart in time are less correlated than closer ones.[1]

Now suppose that for observation registered at time $t = 1, \dots, n_i$ and for subjects $i = 1, \dots, N$ the $y_{it} = (y_{i1}, \dots, y_{ik})'$ be an outcome variable and x_{it} be a $q \times 1$ vector of covariates with the assumption of dependence within clusters (subjects) and independence between clusters. A marginal approach such as GEE suppose that the marginal mean μ_{ij} is a function of covariates through a link function such as $g(\mu_{ij}) = x_{ij}'\beta$, also the variance of y_{ij} is a function of the mean $var(y_{ij}) = \text{Ø}V(\mu_{ij})$. Since Ø is known as the dispersion parameter so GEE solves the equation below:

$$\sum_{i=1}^{N} \dot{\mu}_i' V_i^{-1}(y_i - \mu_i) = 0 \qquad (1)$$

where:

$\dot{\mu}_i = \partial\mu_i/\partial\beta$ Is an $n_i \times q$ matrix

$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$

where: $R_i(a)$ being the working correlation matrix and A_i being the diagonal matrix of the marginal variances $var(y_{ij})$.[7]

The GEE approach has some advantages. As long as, the linear predictor and link function are correctly specified. GEE estimates consistent parameters even if the working correlation matrix is misspecified. Models for GEE are widely available in many statistical softwares. Also, when the true correlation structure is closely approximated, this model will provide efficient parameter estimates. Disadvantages of this model are like difficulty in assessing the goodness-of-fit tests due to lack of the likelihood ratio test (LRT) (12), multiple root problems associated with estimating functions like the quasi-likelihood function. Moreover, this model is sensitive due to outliers and if the correlation structure is misspecified the parameter estimates are not efficient.[7,13]

QIF METHOD

QIF is widely based on observing that the inverse of the working correlation structures can be assessed by a linear combination of several basis matrices:

$$R^{-1} \approx \sum_{i=0}^{M} \alpha_1 M_1 \qquad (2)$$

Where:

M_0 is the identity matrix, $M_1, \dots, M_k$ are known basis matrices with values 0 or 1, and $a_0, \dots, a_k$ are unknown coefficients.

The equation (2) holds for some common working correlation structures. For example if the working correlation is exchangeable, then $R^{-1} \approx a_0 I + a_1 M_1$ where M_1 be 0 on the diagonal and 1 elsewhere. If it is an AR (1) correlation structure then $R^{-1} \approx a_0^* I + a_1^* M_1^* + a_2^* M_2^*$ that M_1^* has 1 on the sub-diagonal and 0 elsewhere and M_2^* has 1 on the two corners of the diagonal and the basis matrices for unstructured correlation structure are given by Qu et al.[14] In occasions that it is difficult to specify a proper working correlation structure, the method of hybrid working correlation which combines basis matrices from several working correlations can be used.[15] By substituting the equation (2) in (1) we can reach to the linear combination of the elements of the extended score g_N as follow:

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\beta) = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} \dot{\mu}_i' A_i^{-1}(y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^{N} \dot{\mu}_i' A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}}(y_i - \mu_i) \end{pmatrix}$$

As the vector g_N contains more estimating equations than parameters so we cannot solve directly the equation of g_N=0 for estimating of β whereas in this case β is over identified therefore the generalized method of moments can be used by minimizing QIF:

$$Q_n(\beta) = N g_N(\beta) C_N^{-1}(\beta) g_N(\beta) \quad (3)$$

where:

$C_N(\beta) = N^{-1}\sum_{i=1}^{N} g_i(\beta) g_i'(\beta)$ is the sample covariance matrix.

In the equation (3) we can see that it includes just the regression parameters β and the basis matrices from the working correlation structure with no need of estimating the nuisance correlation parameter. So, QIF estimator can be obtained by $\hat{\beta} = argmin_{\beta} Q_N(\beta)$. As described latter, QIF approach does not depend on specifying a suitable estimation of the correlation parameters. Also, it avoids the multiple-root problem.[4]

Above all, QIF has some limitations. Firstly, QIF depends on the availability of the basis matrices for a given correlation structure. Secondly, QIF is established only for four types of working correlation structures: Independence, Exchangeability, AR-1 and Unstructured. Although, these four structures cover most important cases it would be of interest to develop QIF that accommodates flexible correlation structures. Similar to GEE, when missing data are present, QIF works only under missing completely at random (MCAR).[5]

It is noticeable that when the working correlation matrix is defined correctly, the efficiency of QIF and GEE are identical. In addition, the estimation of GEE parameters are the same as the maximum likelihood in the case of the multivariate normal model. But under the misspecification of working correlation matrix, QIF gives more efficient parameters than GEE.[6]

DATA SET EXAMINED

The TLGS is a survey designed to gain comprehensive knowledge and information about health and care in Iran. This family-based study which is an ongoing prospective population-based longitudinal cohort study is being conducted to determine the risk factors for non-communicable diseases among a representative urban population of Tehran.[11] The design of TLGS includes two major components, a cross-sectional prevalence study of cardiovascular disease with associated risk factors and a prospective 20 year follow up in several phases: phase I in 1999-2001, phase II in 2002-2005, and phase III in 2006-2008 at approximately 3.6-year intervals. This study was conducted on 1106 households who participated in the TLGS. This study is approved by the Obesity Research Center.

MEASUREMENTS

Height and weight were measured rather than self-reported. The dependent variable Body Mass Index (BMI) was calculated as weight in kilograms divided by the square of the height in meters (kg/m^2), and subjects were classified into obese (BMI ≥30 kg/m^2) and non-obese (BMI <30 kg/m^2).

Information about the respondents' age was based on their self-reported birth year and subjects' smoking status were stratified into a smoker vs. non-smoker and their educational level were stratified into four groups: 1) Primary, 2) Secondary, 3) high school, 4) University.

Divorced, widowed and those who have not married were coded as 0 (non-spouse) vs. others who lived with their spouse was coded 1 (spouse).

STATISTICAL ANALYSIS

At first, the GEE model which takes into account to the correlated nature of responses, with a logit link and exchangeable working correlation structure was used to estimate the odds of obesity as a function of the age, gender, educational level, smoking and marital status.

Then QIF method was also used to estimate OR_S and 95 percent confidence intervals.

Through the below formula it is shown that we can compare the efficiency of parameter estimates from GEE and QIF:

$$\text{Relative Efficiency (RE)} = \frac{\text{mean squared error of GEE estimator}}{\text{mean squared error of QIF estimator}}$$

If RE>1 so QIF is more efficient than GEE, if RE<1 so GEE is more efficient than QIF and

if RE=1 they give the same results.[5,10,16]

All analyses were carried out by using SPSS and SAS softwares.

RESULTS

The mean BMI of males was 27.66 kg/m^2 (95 percent CI: 27.44-27.89). The females had a mean BMI 29.37 kg/m^2 (95 percent CI: 29.09-29.64).

Results were obtained from fitting Models in GEE and QIF methods (Tables 1, 2). The odds ratio estimates for two methods changed only slightly but the RE of parameter estimates from GEE and QIF was 1.23.

EXPLANATION OF RESULTS (TABLE 2)

❑ Female participants had significantly hig-

TABLE 1: Results of parameter estimates using two methods for a random sample of 3203 participants in Tehran, 1999-2008.

		GEE[a]			QIF[b]	
Covariates	**Estimate**	**SE[c]**	**p-value**	**Estimate**	**SE**	**p-value**
Intercept	-2.091	0.390	<0.001	-2.99	0.295	<0.001
Non-Married	-0.397	0.111	<0.001	-0.443	0.113	<0.001
Educational level						
Secondary	0.132	0.326	0.69	0.126	0.328	0.70
High school	0.28	0.112	0.01	0.222	0.111	0.047
University	0.826	0.147	<0.001	0.840	0.149	<0.001
Smoker	-0.001	0.106	0.99	0.041	0.108	0.86
Male	-0.792	0.095	<0.001	-0.864	0.096	<0.001
Age	0.024	0.003	<0.001	0.025	0.003	<0.001

[a] Generalized estimating equations
[b] Quadratic inference functions
[c] Standard errors

TABLE 2: Adjusted[a] odds ratios for obesity and confidence intervals using QIF[b] and GEE[c] method for the TLGS[d] study.

	GEE		QIF	
Covariates	**Odds ratio**	**95% CI[e]**	**Odds ratio**	**95% CI**
Non-married	0.671*	0.5403 – 0.265	0.641*	0.513 – 1.946
Educational level				
Secondary	1.141	1.661 – 0.835	1.134	0.596 – 1.677
High school	1.327*	1.064 – 2.166	1.249*	1.003 – 1.555
University	2.284*	1.712 – 1.654	2.317*	1.727 – 3.109
Smoker	0.998	0.819 - 3.047	1.042	0.843 – 1.288
Male	0.452*	0.375 – 0.545	0.421*	0.349 – 0.508
Age	1.024*	1.018 – 1.030	1.025*	1.019 – 1.031

[a] Adjusted for all other variables in the table.
[b] Quadratic inference functions
[c] Generalized estimating equations
[d] Tehran Lipid and Glucose Study
[e] Confidence Interval.

her odds of obesity than their male counterparts (OR = 2.37, 95% CI: 1.9624-2.8693).

❑ Age was directly associated with obesity (OR=1.026, 95% CI: 1.080-1.0335). Each 1-year increase in age had at least 8% increases in the odds of obesity in women.

❑ Obesity odds ratio was 0.95 (but non-significantly) for non-smokers (95%CI: 0.7762-1.1858).

❑ Married participant had significantly higher odds of obesity than their non-married counterparts (OR = 1.55, 95% CI: 1.2476-1.9469).

❑ An inverse association observed between educational level and obesity. Using primary as reference group, obesity odds ratios were non-significantly 0.88 (95% CI: 0.4631-1.6775) and significantly, 0.80 (95% CI: 0.6429-0.9970) and 0.43 (95% CI: 0.3216-0.5790) for secondary , high school and high level, respectively.

DISCUSSION

In this study we demonstrated that QIF approach has more efficiency than GEE method. In this paper a comparative study between two methods is conducted via an illustrative example, using data from the TLGS database and it included 1106 households. OR estimates and 95% confidence intervals were calculated using QIF method. Overall, we obtained similar parameter estimates from GEE and QIF methods. Our results indicated the improvement of the QIF method over the GEE method (RE= 1.23). Our results were consistent with the findings by simulation results of Qu et al. and also Odueyungbo et al. who compared GEE and QIF by using data from the National Longitudinal Survey of Children and Youth.[5,10]

In our study, there was a positive association between age and obesity. Our results were consistent with some studies.[17-24]

In most studies, participants with lower educational level were obese than those with higher educational level. Our results were consistent with these studies.[24-28]

We found that non-married participants were less likely to be obese than their married counterparts. Our results are in line with most studies.[22,29-31]

One of the limitations of this study is that the physical activity and economic index were not used in our investigation. Our study had several strengths. It was performed in a nationally representative sample of the Iranian households. This is one of the first studies that compare the efficiency of parameter estimates from QIF and GEE using exchangeable working correlation in clusters (TLGS family study). Height and weight were actually measured rather than self-reported. It is well known that self-reports underestimate the prevalence of obesity.[32-34]

In summary, QIF method is better than GEE in the case of misspecification of the working correlation matrix. Moreover, because of incompetence of GEE, results of RE and better parameter estimation with QIF, it is expected that QIF become a newer approach which will be used more frequently than GEE in the future.

REFERENCES

1. Kleinbaum DG, Klein M. Logistic regression for correlated data. Logistic Regression: A Self Learning Text. 2nd ed. New York: Springer; 2002. p. 327-75.
2. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22.
3. Agresti A. Modeling correlated, clustered responses. An Introduction to Categorical Cata Analysis. 2nded. Hoboken, New Jersey: Wiley & Sons; 2006. p. 277-90.
4. Song PX-K, Jiang Z, Park E, Qu A. Quadratic inference functions in marginal models for longitudinal data. Stat Med 2009;28(29):3683-96.
5. Odueyungbo A, Browne D, Akhtar-Danesh N, Thabane L. Comparison of generalized estimating equations and quadratic inference functions using data from the National Longitudinal Survey of Children and Youth (NLSCY). BMC Med Res Methodol 2008;8:1-10.
6. Preisser JS, Qaqish BF. Robust regression for clustered data with application to binary responses. Biometrics 1999;55(2):574-9.

7. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Generalize estimating equations for longitudinal data analysis. Longitudinal data analysis. Hand Books of Modern Statistical Methods. 1st ed. UK: Taylor & Francis Group; 2009. p.43-78.

8. Wang Y-G, Carey V. Working correlation structure misspecification, estimation and covariate design. Biometrika 2003;90(1):29-41.

9. Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 1991;47(3):825-39.

10. Qu A, Lindsay BG, Li B. Improving generalized estimating equations using quadratic inference functions. Biometrika 2000;87(4): 823-36.

11. Azizi F, Ghanbarian A, Momenan AA, Hadaegh F, Mirmiran P, Hedayati M, et al; Tehran Lipid and Glucose Study Group. Prevention of non-communicable disease in a population in nutrition transition: Tehran Lipid and Glucose Study phase II. Trials 2009:10-5.

12. Hardin JW, Hilbe JM. Generalized estimating equations. Generalized Estimating Equations. 1st ed. Florida: Chapman & Hall/CRC Press; 2002. p.104-6.

13. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Marginal models. Analysis of Longitudinal Data. 2nd ed. UK: Oxford University Press; 2013. p.141-67.

14. Small CG, Wang J, Yang Z. Eliminating multiple root problems in estimation. Statist Sci 2000;15(4):313-41.

15. Qu A, Lindsay B. Building adaptive estimating equations when inverse of covariance estimation is difficult. Journal of the Royal Statistical Society, Series B 2003;65(1):127-42.

16. Qu A, Li R. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics 2006;62(2):379-91.

17. Casella G, Berger RL. Properties of a random sample. Statistical Inference. 2nd ed. California: Duxbury Press; 2002. p.207-25.

18. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States,1999-2004. JAMA 2006;295(13):1549-55.

19. Lewis CE, Jacobs DR Jr, McCreath H, Kiefe CI, Schreiner PJ, Smith DE, et al. Weight gain continues in the 1990s: 10 year trends in weight and overweight from the CARDIA Study. Coronary artery risk development in young adults. Am J Epidemiol 2000;15(12): 1172-81.

20. French SA, Story M, Jeffery RW. Environmental influences on eating and physical activity. Annu Rev Public Health 2001;22: 309-35.

21. Bakhshi E, Eshraghian MR, Mohammad K, Foroushani AR, Zeraati H, Fotouhi A, et al. The positive association between number of children and obesity in Iranian women and men: results from the National Health Survey. BMC Public Health 2008;8:213-8.

22. Bakhshi E, Eshraghian MR, Mohammad K, Foroushani AR, Zeraati H, Fotouhi A, et al. Sociodemographic and smoking associated with obesity in adult women in Iran: results from the National Health Survey. J Public Health (Oxf) 2008;30(4):429-35.

23. Bakhshi E, Mohammad K, Eshraghian MR, Seifi B. Factors related to obesity among Iranian men: results from the National Health Survey. Public Health Nutr 2010;13(9):1389-94.

24. Bakhshi E, Seifi B, Biglarian A, Mohammad K. Factors associated with obesity in Iranian elderly people: results from the national health survey. BMC Res Notes 2011;4:538.

25. Bakhshi E, Seifi B, Biglarian A, Mohammad K. Changes in body mass index across age groups in Iranian Women: results from the National Health Survey. J Nutr Metab 2012: 2012;848403.

26. Sarilo-Lähteenkorva S, Lahelma E. The association of body mass index with social and economic disadvantage in women and men. Int J Epidemiol 1999;28(3):445-9.

27. Molarius A, Seidell JC, Sans S, Tuomilehto J, Kuulasmaa KE. Educational level, relative body weight, and changes in their association over 10 years: an international perspective from the WHO MONICA Project. Am J Public Health 2000;90(8):1260-8.

28. Wamala SP, Wolk A, Orth-Gomér K. Determinants of obesity in relation to socioeconomic status among midlle aged Swedish women. Prev Med 1997;26(5 Pt 1):734-44.

29. Klumbiene J, Petkeviciene J, Helasoja V, Prättälä R, Kasmel A. Sociodemographic and health behaviour factors associated with obesity in adult populations in Estonia, Finland and Lithuania. Eur J Public Health 2004;14(4): 390-4.

30. Jeffery RW, Rick AM. Cross-sectional and longitudinal associations between body mass index and marriage-related factors. Obes Res 2002;10(8):809-15.

31. Sobal J, Rauschenbach B, Frongillo EA. Marital status changes and body weight changes: a US longitudinal analysis. Soc Sci Med 2003;56(7):1543-55.

32. Bakhshi E, Eshraghian MR, Mohammad K, Seifi B. A comparison of two methods for estimating odds ratios: results from the National Health Survey. BMC Med Res Methodol 2008;8:78.

33. Bostrom G, Diderichsen F. Socioeconomic differences in misclassification of height, weight and body mass index based on questionnaire data. Int J Epidemiol 1997;26(4):860-6.

34. Kuskowska-Wolk A, Bergstörm A, Boström G. Relationship between questionnaire data and medical records of height, weight and body mass index. Int J Obes Relat Metab Disord 1992;16(1):1-9.