



Universiteit
Leiden
The Netherlands

Long-term consequences of differences in early growth : epidemiological aspects

Euser, A.M.

Citation

Euser, A. M. (2009, December 8). *Long-term consequences of differences in early growth : epidemiological aspects*. Retrieved from <https://hdl.handle.net/1887/14485>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14485>

Note: To cite this publication please use the final published version (if applicable).

4

Reliability studies can be designed more efficiently by using variance components estimates from different sources

A.M. Euser
S. le Cessie
M.J.J. Finken
J.M. Wit
F.W. Dekker

on behalf of the Dutch POPS-19 Collaborative Study Group

J Clin Epidemiol 2007; 60: 1010-1014

Abstract

Objectives

Reliability studies are frequently organized within the context of a large (multicenter) study, with only a small sample of subjects measured by the observers of the large study. To estimate interobserver reliability, data from the large study are not frequently used. In this article, the advantages of combining data from the reliability study and the large study to improve the estimation of intra-class correlation coefficients (ICCs) are highlighted.

Study design and setting

This was done within the scope of estimating fat percentages in the Project On Preterm and Small-for-gestational-age infants-19 (POPS-19) study and with simulations. To calculate ICCs, three approaches were used: (1) the classical approach using data from a reliability study only, (2) the combined variances approach using inter-subject variances from the POPS-19 study, and (3) the maximum likelihood approach using all data.

Results

The ICCs (95% confidence interval [CI]) for fat percentage calculated by the three approaches were 0.84 (0.57, 0.99), 0.94 (0.90, 0.97), and 0.94 (0.88, 0.97), respectively.

Conclusion

The efficient use of data by combining data from a small reliability study with the data from the large study itself for the calculation of ICCs will lead to more precise ICCs.

Introduction

The reliability of clinical measurements is an important issue in the design and interpretation of studies. A high degree of measurement error resulting in a poor reliability generally leads to underestimation of the strength of the associations studied. This affects the interpretation of results and can even lead to erroneous negative conclusions. To compensate for a lack of precision, sometimes a higher number of subjects is included or more repeated measurements per subject are obtained, but both are associated with less efficiency or more costs. In many large and often multicenter studies, reliability of clinical measurements can even be lower because multiple observers are involved in data collection.

Therefore, in large (multicenter) studies reliability of clinical measurements is sometimes assessed in special substudies to enhance the interpretation of the results of the main study. For example, Visser et al. studied the reliability of the Subjective Global Assessment of nutritional status¹ in a small substudy of the large multicenter NECOSAD cohort on risk factors for mortality in dialysis patients² whereas Klipstein-Grobusch et al. described the reliability of anthropometric measurements³ assessed in the European Prospective Investigation into Cancer-Potsdam Study Cohort.⁴ The classical approach in this situation is to perform a reliability study with a small, random sample of about at least 10 study participants (subjects) to be measured by all observers involved in the large study.⁵ From the measurements in these reliability studies, indicators of reliability, for example intra-class correlation coefficients (ICCs), can be estimated.⁶ In this approach, only data from the subjects participating in the reliability study are used, whereas the data from the large study are not used.

There are several methods to estimate ICCs by combining data from the reliability study and the large study. The first approach is to determine the interobserver and error variance in the reliability study and the inter-subject variance in the large study itself. This is in the line of Streiner and Norman who describe a formula to apply a known ICC to a different, more heterogenous population.⁷ The second approach is to combine the data from both studies with maximum likelihood (ML) methods.

As these approaches are not frequently used in health science literature, the aim of this article was to highlight the advantages of combining data from a reliability study with data from the large study itself for the calculation of ICCs. For this purpose, we apply existing statistics to a novel context. The ICCs calculated in this way will be more accurate because of using data directly from the population of interest and also more precise because of the larger study population used. We will show this in a data set of the Project On Preterm and Small-for-gestational-age infants-19 (POPS-19) study,⁸ and with simulations.

Subjects and methods

Study population

The POPS-19 study is a Dutch national prospective cohort study in young adults aged 19 years born before 32 weeks' gestation. Among other measurements, skinfold thickness was measured at four regions to determine fat mass and fat distribution in 403 19 years old. The POPS-19 study was organized as a multicenter study with 10 research centers all over the country and 15 observers.⁸ When the POPS-19 study was started, a reliability study was organized with four healthy young adults who had their skinfolds measured in all 10 research centers by 13 out of the 15 POPS-19 observers. Due to practical circumstances (limited space in a small car for four people to travel about the whole country for various measurements, including time consuming ones), this reliability study only had a small sample size. In both studies, skinfold thickness measurements were performed in duplicate on the left side of the body at the triceps, biceps, subscapular, and iliacal regions. In the final analyses, the mean of each duplicate measurement was used. Fat percentage was computed from the sum of the four skinfolds.⁹ The study was approved by the medical ethics committee of all participating centers, and written informed consent was obtained from all participants.

Statistical analysis

Suppose that there are n subjects in the large study where each subject is measured by one observer. Furthermore, let J be the number of observers and I be the number of subjects in the reliability study. For some given variable X , we denote by X_{ij} the measurement of the j th observer made on the i th subject for $i = 1, \dots, I, j = 1, \dots, J$. We modeled the data as

$$Z_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

where μ is some fixed parameter, and where α_i the subject effect, β_j the observer effect and ε_{ij} are independent random effects, normally distributed with mean 0 and with between-subject variance σ_S^2 ; interobserver variance σ_O^2 ; and error variance σ_E^2 ; respectively.

Interobserver reliability was measured with:

$$\text{ICC}_{\text{inter}} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_O^2 + \sigma_E^2} \quad (2)$$

We consider three different approaches to estimate the variance components σ_S^2 ; σ_O^2 ; and σ_E^2 and $\text{ICC}_{\text{inter}}$. For all approaches logarithmical transformations of the skinfold measurements were performed because of the skewed distribution of errors of these variables.

Classical approach

In this approach, all variance components are estimated using only the data from the reliability study. The design of the reliability study is balanced (all subjects are measured by all observers) and variance components can be estimated using classical analysis of variance, which yields ICCs according to equation (2) and confidence intervals (CIs). See for details Shrout and Fleiss (formula 2, 1): ICC with random observer effect, single ratings.⁶

Combined variances approach

Here, σ^2_{O} and σ^2_{E} are estimated from the reliability study, and the inter-subject variability σ^2_{S} from the multi-center data. To estimate the inter-subject variability, we estimated the total variance σ^2_{T} of the variable X by the variance from the data of the large study. We assumed that $\sigma^2_{\text{S}} = \sigma^2_{\text{T}} - \sigma^2_{\text{O}} - \sigma^2_{\text{E}}$. Subsequently, estimates of ICCs were obtained by plugging-in estimates of variance components in equation (2). Ninety-five percent CIs of these ICCs can be obtained using the delta method (details are given in Appendix A), but is not straightforward to carry out because an estimate of the covariance matrix of the estimated variance components is needed.

ML approach

Both data sets are pooled and ML methods are used. Combining the data of both studies yields a data set with $(n + l)$ subjects, where some of the subjects in this data set are measured by all observers, others by only one. In fact, one can see this as a very large reliability study with many missing observations (because not all subjects are measured by all observers). In this design, variance components can be estimated using ML or restricted maximum likelihood (REML). We used REML, as the REML estimator is known to be in general less biased than the ML estimator,¹⁰ (page 66–69). This can be carried out with software for linear mixed models like SAS PROC mixed (SAS institute Inc., Cary, NC, USA). This yields estimates of σ^2_{S} ; σ^2_{O} ; and σ^2_{E} and of the covariance matrix of the estimates. The ICC is calculated by plugging these estimates into equation (2). Again 95% CI can be obtained using the delta method (see Appendix A).

In the Section 3, the three different approaches to the estimation of ICCs will be applied on the POPS-19 data. We also compare the efficiency of the different approaches in a simulation study using SAS version 8.2 (SAS institute Inc., Cary, NC, USA). We simulated data from model (1), with mean $\mu=0$, and with variance of the subject, observer and residual effect equal to $\sigma^2_{\text{S}}=8$, $\sigma^2_{\text{O}}=1$, and $\sigma^2_{\text{E}}=1$, respectively. This implies that the ICC = 0.80. The parameter values were based upon the values of the triceps skinfold in the study example mentioned above, in which ICCs were all around 0.80 with quite similar observer and residual variances. Based on the POPS-19 example, we assumed a small reliability study with 4 subjects measured by 10 observers and a large study with 400 subjects each measured by only one

observer. Data were simulated 1,000 times from this setup. Because of the small number of subjects in the POPS-19 reliability study, we also repeated simulations with 10, 25, and 50 subjects in the reliability study.

Results

General characteristics of the subjects from the POPS-19 study and the reliability study are displayed in Table 1. On average, the four subjects of the reliability study were somewhat older, and had greater body mass index (BMI) and sum of skinfolds than the POPS-19 participants. The anthropometric characteristics of all four subjects were well in the range of the POPS-19 participants.

Table 1. Characteristics of the study participants; means (SD)

Characteristics	POPS-19 study ($n = 403$)	Reliability study ($n = 4$)
Sex (% male)	46.4	50.0
Age	19.3 (0.18)	24.6 (3.6)
(Min, Max)	(19.1, 20.0)	(22.1, 30.0)
BMI (kg/m²)		
Males	21.7 (3.1)	25.3 (1.5)
(Min, Max)	(14.8, 34.7)	(24.3, 26.4)
Females	21.8 (3.4)	21.2 (1.5)
(Min, Max)	(15.6, 38.9)	(20.2, 22.3)
Sum of four skinfolds (mm)		
Males	41.2 (20.5)	81.2 (23.6)
(Min, Max)	(16.0, 130.5)	(64.6, 97.9)
Females	62.2 (22.6)	70.9 (0.57)
(Min, Max)	(7.3, 149.0)	(70.5, 71.3)

The variance components, ICCs, and 95% CIs of the triceps skinfold and fat percentage are presented in Table 2. Due to the small estimated between-subject variance in the reliability study, the classical approach yields lower ICCs compared with the two other approaches. This effect is more pronounced for the triceps skinfold than for the fat percentage. Both with the combined variances approach and with the ML approach, the obtained 95% CIs are much smaller than estimated with the classical approach. The ML approach yields larger estimates of both the between observer and measurement error variance with a slightly larger estimated 95% CI. The other skinfold measurements and derived estimates of body composition showed comparable results with regard to the differences between the various approaches (data not shown).

Table 2. Variance components and ICCs estimated in the POPS-19 data according to the various approaches

Outcome measure and method	σ^2_S	σ^2_O	σ^2_E	ICC	95% CI
Triceps skinfold thickness ^a					
Classical approach	0.00675	0.00249	0.00299	0.55	0.24 - 0.95
Combined variances approach	0.0442	0.00249	0.00299	0.89	0.82 - 0.93
REML approach	0.04151	0.00391	0.00354	0.85	0.74 - 0.91
Fat percentage					
Classical approach	22.0	2.85	1.48	0.84	0.57 - 0.99
Combined variances approach	70.9	2.85	1.48	0.94	0.90 - 0.97
REML approach	71.8	3.27	1.67	0.94	0.88 - 0.97

a For all approaches, logarithmical transformations were performed because of the skewed distribution of errors of the variables.

The variance components, ICCs, and 95% CIs of the triceps skinfold and fat percentage are presented in Table 2. Due to the small estimated between-subject variance in the reliability study, the classical approach yields lower ICCs compared with the two other approaches. This effect is more pronounced for the triceps skinfold than for the fat percentage. Both with the combined variances approach and with the ML approach, the obtained 95% CIs are much smaller than estimated with the classical approach. The ML approach yields larger estimates of both the between observer and measurement error variance with a slightly larger estimated 95% CI. The other skinfold measurements and derived estimates of body composition showed comparable results with regard to the differences between the various approaches (data not shown).

Table 3. Results of 1,000 simulations to compare the three estimation approaches with different numbers of subjects in the reliability study

Number of subjects in the reliability study	Estimation approach	σ^2_S	σ^2_O	σ^2_E	ICC
		True parameters	8	1	1
4 (POPS-19 example)	1. Classical approach	6.34 (0.99; 20.95)	0.91 (0.19; 2.17)	0.98 (0.62; 1.49)	0.76 (0.32; 0.92)
	2. Combined variances approach	7.89 (6.52; 9.25)	0.91 (0.19; 2.17)	0.98 (0.62; 1.49)	0.81 (0.69; 0.88)
	3. REML approach	8.01 (6.84; 9.10)	0.94 (0.27; 2.01)	0.99 (0.63; 1.46)	0.80 (0.71; 0.86)
10	1. Classical approach	7.42 (2.75; 14.73)	0.94 (0.29; 1.93)	0.99 (0.74; 1.27)	0.79 (0.57; 0.89)
	2. Combined variances approach	7.90 (6.69; 9.13)	0.94 (0.29; 1.93)	0.99 (0.74; 1.27)	0.80 (0.71; 0.87)
	3. REML approach	8.00 (6.97; 9.06)	0.94 (0.32; 1.91)	0.99 (0.75; 1.27)	0.80 (0.73; 0.87)
25	1. Classical approach	7.67 (4.69; 12.11)	0.91 (0.35; 1.96)	0.99 (0.86; 1.16)	0.80 (0.68; 0.88)
	2. Combined variances approach	7.89 (6.76; 9.07)	0.91 (0.35; 1.96)	0.99 (0.86; 1.16)	0.80 (0.72; 0.86)
	3. REML approach	7.96 (6.98; 9.00)	0.91 (0.35; 2.00)	0.99 (0.86; 1.16)	0.81 (0.73; 0.86)
50	1. Classical approach	7.83 (5.45; 10.50)	0.89 (0.35; 1.96)	1.00 (0.89; 1.11)	0.80 (0.70; 0.87)
	2. Combined variances approach	7.86 (6.77; 9.08)	0.89 (0.35; 1.96)	1.00 (0.89; 1.11)	0.81 (0.71; 0.86)
	3. REML approach	8.00 (6.99; 8.93)	0.90 (0.34; 1.97)	1.00 (0.89; 1.11)	0.81 (0.72; 0.86)

The median and (between brackets) 5th and 95th percentiles of the 1,000 estimates are given.

The results of the simulation analyses are summarized in Table 3. When the number of subjects is small ($l = 4$), the 90% ranges of the ICCs from the combined variances and the REML approach are much smaller than that of the ICC estimated with the classical approach. The REML approach is a little more precise with a smaller 90% range compared to the combined variances approach. The median value of the estimated inter-subject variance as shown in Table 3 was 6.34, much smaller than the true value of 8.00. The median estimates of the ICCs were all close to the true value of 0.80.

With increasing numbers of subjects measured, the median of the estimated inter-subject variances comes closer to the true parameter 8.00. For both 10 and 25 subjects, the ICCs estimated with the combined variance approach and the REML approach still have a considerably smaller 90% range than those estimated with the classical approach. With 50 subjects, this effect is less pronounced. These simulations show that using the REML approach with 10 subjects in the reliability study (ICC, 90%; range, 0.73–0.87) is at least as precise as measuring 50 subjects and using the classical approach (ICC, 90%; range, 0.70–0.87).

Discussion

In this article, two approaches are described which improve the precision of the estimation of ICCs in the context of a reliability study organized within a large study. These approaches were compared with the classical approach, that is, estimating all variance components in the small reliability study. With a relatively simple method, the inter-subject variance is estimated in the large study itself, whereas the other variance components are estimated in a reliability study. The other method, which is somewhat more precise, uses ML on the combined data from both studies.

The advantage of these approaches is that they obviate two possible shortcomings of the estimation of ICCs according to the classical approach. Firstly, due to chance the subjects in the reliability study might not form a representative sample of the subjects in the large study with a different inter-subject variance. This situation in which an ICC is applied to a different, more heterogeneous population has been described by Streiner and Norman⁷ page 147, and before by Lord and Novick,¹¹ page 130. In combined variance and ML approaches data of the population of interest, namely the large study, are used to estimate the inter-subject variability, circumventing this problem. Secondly, a relatively small number of subjects is used in the classical approach, whereas with our approaches in which all available data are used a more precise estimation can be carried out with smaller CIs as a result.

A limitation of our study is the small sample size of four subjects we had to use in the reliability study due to practical circumstances. We assume that measurement error in skinfold thicknesses is not remarkably different in 19 and 20–30 years old, but still the small size could have influenced the value of the inter-subject variance found. In contrast, the interobserver variance was based on measurements of 13 observers. In the simulations, it can be seen that the estimated inter-subject variance with four subjects in the reliability study could differ much from the real parameter, which shows the advantages of our approaches. We studied the generalizability of our results by repeating simulations with larger numbers of subjects as commonly used in reliability studies^{3, 5, 12} and.¹³ With 10 or 25 subjects, the inter-subject variance and ICC did not differ much between all approaches, but the combined variances approach and the REML approach are still preferable to the classical approach regarding the precision of the estimated ICC as reflected in the smaller 90% ranges.

For clarity, in this article we used the means of duplicate measurements, and modeled the data as $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$. However, the described approaches of calculating ICCs can also be extended to a model using the separate duplicate measurements on a subject, subdividing the error variance into variance due to observer–subject variance and residual error variance. This will give comparable results.

In conclusion, we have shown the value of our novel approaches to estimate more precise ICCs with the efficient use of combined data in the POPS-19 study and we suggest that this approach can also be used in other studies concerning the reliability of outcomes in a large study. It is important to have precise information about the interobserver reliability of the outcome measurements, because this will influence the associations found between determinant and outcome in the large study. Low reliability will give noise and dilution, or even confounding of the associations found. With our approaches more precise estimations of ICCs are obtained, and we suggest to take this innovation into account when designing future reliability studies in the context of a large study.

Acknowledgments

The POPS study at 19 years of age was supported by grants from the Netherlands Organisation for Health Research and Development (ZonMw), Edgar Doncker Foundation, Foundation for Public Health Fundraising Campaigns, Phelps Foundation, Swart-van Essen Foundation, Foundation for Children's Welfare Stamps, TNO Prevention and Health, Netherlands Organisation for Scientific Research (NWO), Dutch Kidney Foundation, Sophia Foundation for Medical Research, Stichting Astmabestrijding, Royal Effatha Guyot group. The cooperation of A. Inderson and A.G. van Montfoort is gratefully acknowledged.

Appendix A

In this appendix, it is shown how a confidence interval can be calculated for an intra class correlation, using the delta method.

The intra class correlation is defined as:

$$ICC = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_O^2 + \sigma_E^2}$$

In the maximum likelihood approach, the variance components are estimated by either ML or REML, which also yields an estimate of the covariance matrix of the estimates.

To obtain a more closely normally distributed variate, the ICC is transformed using the Fisher-Z transformation:

$$z = \log\left(\frac{1+ICC}{1-ICC}\right) = \log\left(\frac{2\sigma_S^2 + \sigma_O^2 + \sigma_E^2}{\sigma_O^2 + \sigma_E^2}\right)$$

For convenience, we use matrix notations. Let τ be the vector of variance components: $\tau = (\sigma_S^2 \ \sigma_O^2 \ \sigma_E^2)^T$, where the superscript T indicates a transposed vector. The delta method gives that

$$\text{var}(\hat{Z}) \approx \left(\frac{\partial Z}{\partial \tau}\right)^T \text{var}(\hat{\tau}) \frac{\partial Z}{\partial \tau}, \quad (*)$$

with $\text{var}(\hat{\tau})$ the covariance matrix of the estimated variance components. It is straightforward to show that:

$$\frac{\partial Z}{\partial \tau} = \frac{1}{2\sigma_S^2 + \sigma_O^2 + \sigma_E^2} \begin{pmatrix} 2 \\ -2\sigma_S^2/(\sigma_O^2 + \sigma_E^2) \\ -2\sigma_S^2/(\sigma_O^2 + \sigma_E^2) \end{pmatrix}$$

and this can be plugged in (*) to obtain the variance of \hat{Z} . A 95% confidence interval for Z can be calculated by:

$$(\hat{Z} - 1.96\sqrt{\text{var}(\hat{Z})} \ \hat{Z} + 1.96\sqrt{\text{var}(\hat{Z})})$$

This interval can be transformed back to an interval for ICC. If lwb and upb are respectively the lower and upper bound of the 95% CI for Z, the 95% CI for the ICC is given by:

$$\left(\left(\frac{e^{lwb} - 1}{e^{lwb} + 1} \right), \left(\frac{e^{upb} - 1}{e^{upb} + 1} \right) \right).$$

In the combined variances approach, the total variance σ_T^2 was estimated from the large study, while independently σ_O^2 and σ_E^2 were obtained from the reliability study. By writing

$ICC = \frac{\sigma_T^2 - (\sigma_O^2 + \sigma_E^2)}{\sigma_T^2}$, the delta method can be applied in the same way to obtain confidence

intervals in this situation.

Appendix B Supplementary material

Participants of the Dutch POPS-19 Collaborative Study Group

TNO Prevention and Health, Leiden (E.T.M. Hille, C.H. de Groot, H. Kloosterboer-Boerrigter, A.L. den Ouden, A. Rijpstra, S.P. Verloove-Vanhorick, J.A. Vogelaar); Emma Children's Hospital AMC, Amsterdam (J.H. Kok, A. Ilsen, M. van der Lans, W.J.C. Boelen-van der Loo, T. Lundqvist, H.S.A. Heymans); University Hospital Groningen, Beatrix Children's Hospital, Groningen (E.J. Duiverman, W.B. Geven, M.L. Duiverman, L.I. Geven, E.J.L.E. Vrijlandt); University Hospital Maastricht, Maastricht (A.L.M. Mulder, A. Gerver); University Medical Center St Radboud, Nijmegen (L.A.A. Kollée, L. Reijmers, R. Sonnemans); Leiden University Medical Center, Leiden (J.M. Wit, F.W. Dekker, M.J.J. Finken); Erasmus MC – Sophia Children's Hospital, University Medical Center Rotterdam (N. Weisglas-Kuperus, M.G. Keijzer-Veen, A.J. van der Heijden, J.B. van Goudoever); VU University Medical Center, Amsterdam (M.M. van Weissenbruch, A. Cranendonk, H.A. Delemarre-van de Waal, L. de Groot, J.F. Samsom); Wilhelmina Children's Hospital, UMC, Utrecht (L.S. de Vries, K.J. Rademaker, E. Moerman, M. Voogtsgeerd); Máxima Medical Center, Veldhoven (M.J.K. de Kleine, P. Andriessen, C.C.M. Dielissen-van Helvoirt, I. Mohamed); Isala Clinics, Zwolle (H.L.M. van Straaten, W. Baerts, G.W. Veneklaas Slots-Kloosterboer, E.M.J. Tuller-Pikkemaat); Royal Effatha Guyot Group, Zoetermeer (M.H. Ens-Dokkum); and Association for Parents of Premature Babies (G.J. van Steenbrugge).

References

1. Visser R, Dekker FW, Boeschoten EW, Stevens P, Krediet RT. Reliability of the 7-point subjective global assessment scale in assessing nutritional status of dialysis patients. *Adv Perit Dial* 1999; 15:222-225.
2. Jansen MA, Korevaar JC, Dekker FW, Jager KJ, Boeschoten EW, Krediet RT. Renal function and nutritional status at the start of chronic dialysis treatment. *J Am Soc Nephrol* 2001; 12:157-163.
3. Klipstein-Grobusch K, Georg T, Boeing H. Interviewer variability in anthropometric measurements and estimates of body composition. *Int J Epidemiol* 1997; 26 Suppl 1:S174-S180.
4. Heidemann C, Hoffmann K, Spranger J, Klipstein-Grobusch K, Mohlig M, Pfeiffer AFH et al. A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC) - Potsdam Study cohort. *Diabetologia* 2005; 48:1126-1134.
5. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition* 1999; 82:165-177.
6. Shrout PE, Fleiss JL. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychological Bulletin* 1979; 86:420-428.
7. Streiner DL, Norman GR. Reliability. In: Streiner DL, Norman GR, editors. *Health measurement scales - a practical guide to their development and use*. 3 ed. Oxford: Oxford University Press; 2003. 126-152.
8. Euser AM, Finken MJ, Keijzer-Veen MG, Hille ET, Wit JM, Dekker FW. Associations between prenatal and infancy weight gain and BMI, fat mass, and fat distribution in young adulthood: a prospective cohort study in males and females born very preterm. *Am J Clin Nutr* 2005; 81:480-487.
9. Durnin JVG, Rahaman MM. Assessment of Amount of Fat in Human Body from Measurements of Skinfold Thickness. *British Journal of Nutrition* 1967; 21:681-&.
10. Diggle P.J., Heagerty P, Liang K, Zeger S.L. *Analysis of Longitudinal Data*. 2 ed. Oxford: Oxford university press; 2002. 66-69.
11. Lord F.M, Novick M.R. *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley publishing company; 1968. 130.
12. Rouwmaat PHM, Everaert D, Stappaerts KH, Aufdemkampe G. Reliability of manual skinfold tests in a healthy male population. *Journal of Manipulative and Physiological Therapeutics* 1998; 21:327-332.
13. Moreno LA, Joyanes M, Mesana MI, Gonzalez-Gross M, Gil CM, Sarria A et al. Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents. *Nutrition* 2003; 19:481-486

