

Reliability of period of gestation determined by ultrasound scan measurements

Chrishantha Abeysena, Pushpa Jayawardana

International Journal of Collaborative Research on Internal Medicine & Public Health Vol. 3 No. 5 (May 2011)

International Journal of Collaborative Research on Internal Medicine & Public Health (IJCRIMPH)

ISSN 1840-4529 | Journal Type: Open Access | Volume 3 Number 5

Journal details including published articles and guidelines for authors can be found at: http://www.iomcworld.com/ijcrimph/

To cite this Article: Abeysena C, Jayawardana P. Reliability of period of gestation determined by ultrasound scan measurements. *International Journal of Collaborative Research on Internal Medicine & Public Health.* 2011; 3(5):334-339.

Article URL: http://iomcworld.com/ijcrimph/ijcrimph-v03-n05-01.htm

Correspondence concerning this article should be addressed to Dr. Chrishantha Abeysena; Senior Lecturer / Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka / Tel: [University] 0094 11 2953411 / Fax: [University] 0094 11 2958337 / Email: chrishantha-abeysena@mfac.kln.ac.lk

Paper publication: 30 May 2011

International Journal of Collaborative Research on Internal Medicine & Public Health

Editors-in-Chief: Asst. Prof. Dr. Jaspreet S. Brar (University of Pittsburgh, USA) Forouzan Bayat Nejad

Executive Editor: Mostafa Nejati

Deputy Editor: Dr. Mensura Kudumovic (University of Sarajevo, Bosnia & Herzegovina)

Associate Editors: Dr. Monica Gaidhane Dr. Suresh Vatsyayann (FreeGP, New Zealand)

Reliability of period of gestation determined by ultrasound scan measurements

Chrishantha Abeysena (1)^{*}, Pushpa Jayawardana (2)

Senior Lecturer / Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka
Professor / Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

* Corresponding author; Email: chrishantha-abeysena@mfac.kln.ac.lk

ABSTRACT

Objective: To determine reliability of period of gestation determined by three independent raters using four different foetal measurements.

Methods: One hundred and eighty pregnant women were divided into three equal groups. Each group was assigned a rater to perform ultrasound scan to measure bi-parietal diameter, femur length, abdominal circumference and head circumference and to compute the respective periods of gestation using these four measurements. Reliability between periods of gestation derived by each rater from above four measurements were analysed using repeated measure ANOVA. Results were expressed as intra-class correlation coefficients (ICCs) and coefficients of variation (CsOV).

Results: For Raters I (F= 6.47; p=0.001) and II, (F= 4.80; p= 0.003), computations using abdominal circumferences resulted in the lowest mean periods of gestation (PsOG). For Rater III, computations using both femur length and abdominal circumference resulted in the lowest mean periods of gestation (F= 7.5; p=0.001). ICCs were 0.73 (95%CI 0.64–0.81) for Rater I, 0.78 (95% CI 0.70–0.85) for Rater II and 0.87 (95% CI 0.81–0.91) for Rater III

When comparing CsOV, the highest variation for Raters I and III was observed for femur length. For Rater II it was bi-parietal diameter. The lowest variation for Rater I was observed for head circumference and for Raters II and III for abdominal circumference. The highest CsOV of all the PsOG were demonstrated by Rater III.

When comparing the differences between the highest and the lowest values for each period of gestation determined, the difference was more than two weeks for 38% (n=23), 24% (n=14) and 22% (n=13) of observations made by Raters I, II and III respectively.

Conclusions: Reliability of period of gestation depends on the type of measurement taken, method of assessment and the rater who performs the measurements. Our findings are not conclusive enough to recommend any PsOG based on specific measurement more reliable than others. In-service training of the obstetricians is likely to improve the reliability of PsOG determined using ultra sound scan measurements.

Keywords: Gestational age, pregnancy duration, reliability, ultrasound scan

Introduction

Gestational age is usually determined by the date of the woman's last menstrual period. Sometimes a woman may be uncertain of the date of her last menstrual period. Ultrasound scans offer an alternative method for estimating gestational age.¹ It is currently considered to be a safe, non-invasive, accurate and cost-effective investigation of the foetus and there is no strong evidence to suggest that ultrasound harms babies.² It has progressively become an indispensable obstetric tool and plays an important role in the care of every pregnant woman.

The main uses of ultrasonography are determination of gestational age and assessment of foetal size. Foetal body measurements reflect the gestational age of the foetus. This is particularly true in early gestation. In patients with uncertain last menstrual periods, such measurements must be made as early as possible in pregnancy to arrive at a correct dating for the patient. In the latter part of pregnancy measuring body parameters will allow assessment of the size and growth of the foetus and will greatly assist the diagnosis and management in of intrauterine growth retardation.

The most accurate measurement for dating is the crown-rump length of the foetus, which can be done between 7 and 13 weeks of gestation. After 13 weeks of gestation, the foetal age may be estimated by the biparietal diameter (the transverse diameter of the head), the head circumference and the length of the femur (the longest bone in the body).³ The abdominal circumference of the foetus may also be measured and this gives an estimate of the weight and size of the fetus⁴ as well.

Reliability refers to the consistency of a measure. A test is considered reliable if we obtain the same result on repetition of the measurement. Even though it is impossible to

calculate it exactly, there are different ways to estimate it.⁵ Several types of reliability are explained.^{6,7} To assess test-retest reliability, the measurement is taken twice at two different points in time. To assess inter-rater reliability, two or more independent observers have to take the measurement from the same individual and then determine the consistency of the raters estimates. Intra-rater reliability requires taking two or more independent measurements by an individual at two different points in time. Alternative-form reliability uses different methods to measure the same variable. Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. Reliability is an essential component of validity but, on its own, is not a sufficient measure of validity. A test can be reliable but not valid, whereas a test cannot be valid and yet to be unreliable.⁷

Reliability and validity of the period of gestation determined by ultrasound scan depends on several factors: the type of measurement performed, the training and skills of the person performing the scan, duration of gestation for which the scan is performed and the lie of the foetus.⁸ The objective of the study was to determine reliability of the periods of gestation determined using four different foetal measurements.

Methods

Descriptive study was conducted using 180 pregnant women who were recruited before 16th weeks of gestation. They were invited for the ultrasound scan before the completion of the 20th week of gestation. All of them were randomly divided into three equal groups. Each group was assigned a rater who was a Consultant Obstetrician. All women were

subjected to the scan by the respective raters during which bi-parietal diameter, femur length and abdominal and head circumferences of the foetus were measured. The period of gestation (PsOG) derived from each of these measurements and displayed on the machine were noted in the relevant record sheet. All scans were conducted using the same ultrasound machine.

Repeated measure ANOVA was applied to compare the mean values of the different measurements of each rater. Reliability was assessed by computing intra-class correlation coefficients (ICCs) and coefficients of variation (CsOV). The difference between the highest and the lowest values of PsOG derived were computed for each individual mother to determine the proportion of observations that exceeded a difference of two weeks.

Ethical clearance was obtained from the Ethics Review Committee of the Faculty of Medicine, University of Kelaniya.

Results

Statistically significant differences were observed between the mean PsOG computed using the four measurements by all the three raters: Rater I (F= 6.47; p=0.001), Rater II (F= 4.80; p=0.003) and Rater III (F= 7.5; p=0.001) (Tables 1, 2 and 3). For Raters I and II. computations using abdominal circumferences resulted in the lowest mean PsOG. For Rater III, computations using both femur length and abdominal circumference resulted in the lowest mean PsOG Intra class correlation coefficients between PsOG derived taking the four measurements by each rater were 0.73 (95%CI 0.64–0.81) for Rater I, 0.78 (95% CI 0.70-0.85) for Rater II and 0.87 (95% CI 0.81-0.91) for Rater III (Tables 1, 2 and 3).

With regard to CsOV, the highest variation was observed for the PsOG derived using femur length for Raters I and III and for the bi-parietal diameter for Rater II (Tables 1, 2 and 3). The lowest variation was observed for PsOG derived using head circumference by Rater I and for abdominal circumference by Raters II and III. All the four PsOG derived by Rater III recorded the highest CsOV.

When comparing the differences between the highest and the lowest values of each period of gestation determined, a difference of more than two weeks was observed for 38% (n=23), 24% (n=14) and 22% (n=13) of assessments made by Raters I, II and III respectively (Table 4).

Discussion

Periods of gestation based on measurements taken by Rater III had higher reliability in terms of ICC than Raters I and II. Further, highest reliability was for Rater Ш considering the differences between the highest and the lowest PsOG for each individual mother. However in terms of PsOG based on individual four measurements, reliability was lowest for Rater III as all the CsOV were higher than that of Raters I and II. Comparing the CsOV values across the three raters, period of gestation derived from abdominal circumference had the lowest CsOV indicating highest reliability.

Our design mainly assessed alternative-form reliability, where different measurements were used to assess the PsOG in the same woman. This was conducted using three raters even though the use of a single rater would have sufficed and we calculated ICC for periods of gestation separately for each rater. Even though we selected three groups randomly for each rater, there is a possibility that the average true period of gestation for each group might vary. Therefore pooling the data obtained from each rater could bias the ICC as well as making the calculated CsOV uninterpretable. Further, assessment of interrater reliability was not our objective.

The PsOG are derived automatically based on regression equations fed to the ultrasound scan machine. Thus the variations of our results may also be based on the validity of the regression equations which was not under our control. Further, for clinical practice, it is the reliability of PsOG which are more important than foetal body measurements from which they are derived using the regression equations. Hence our focus was on the reliability of the former and not the latter.

Reliability and validity are often confused, but the terms actually describe two completely different concepts, although they are often closely inter-related. Reliability is an essential component of validity but, on its own, is not a sufficient measure of validity. A test can be reliable but not valid, whereas a test cannot be valid yet unreliable. We did not assess the validity of the period of gestation. Even though there are several methods of assessing reliability statistically, it is important to applicability clinically. review its We considered a difference of more than two weeks between the highest and the lowest PsOG as clinically significant.

Olesen⁹ pointed out that the gestational ages estimated using ultrasound scanning were 2-3 days shorter than gestational ages estimated by the menstrual dates. However, they have observed a good concordance between the self-reported and ultrasound-estimated gestational ages.⁹ Kieler¹⁰ also compared the expected date of delivery derived from biparietal diameter and the last regular menstrual period and found that 91.8% delivered within +14 days and 61.8% within seven days of expected date of delivery (EDD) derived using the bi-parietal diameter. Corresponding figures were 91.6% and 61.1% respectively for EDD derived from last regular menstrual period.¹⁰ However, comparison of EDD calculated from bi-parietal diameter and last regular menstrual period in the same woman showed that measurements derived using bi-parietal diameter postponed EDD by more than seven days in 18.0% and advanced it more than by seven days in 1.8%.¹⁰ Campbell had reported that bi-parietal diameter measurements done between 12 and 18 weeks gestation were significantly more accurate in gestational predictions (89.4%) than those based on menstrual history.¹ Persson also commented that gestational age estimated using by bi-parietal diameter gave the best reliability, with a standard deviation from true gestational age of 3.2 days.¹² Of the patients with an optimal menstrual history 84.7% delivered within two weeks of the predicted date.¹² The maximum difference between gestational age estimated by biparietal diameter and by femur length had been seven days.¹²

According to a study done by O'Brien femur length had been reported to provide a reproducible determination of length of the fetus.¹³ Johnsen reported that gestational age assessment based on femur length was an equally robust method as using head circumference.¹⁴

In conclusion reliability of PsOG was dependant on the type of measurement taken from the foetus from which they were derived, the method of assessment of reliability and the obstetrician who performed the measurements. Our findings are not conclusive enough to recommend any PsOG based on specific measurement more reliable than others. Inservice training of the obstetricians is likely to improve the reliability of assessments.

References

1. Gardosi J, Geirsson RT. Routine ultrasound is the method of Choice for dating Pregnancy. British Journal of Obstetrics and Gynecology.1998;105: 933-936.

2. Newnham J, Evans SF, Michael CA et al. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. Lancet. 1993;342:887-91.

3. Chudleigh P, Pearce JM. Obstetric ultrasound. Second Edition, Churchill Livingstone. 1992;77-94.

4. Woo, Joseph (2006). "Why and when is Ultrasound used in Pregnancy?". *Obstetric Ultrasound: A Comprehensive Guide*. http://www.ob-ultrasound.net/. Retrieved 2010-02-12.

5. Calculations for reliability. Available http://sportsci.org/resource/stats/relycalc.html. Retrieved 2010-02-12.

6. Litwin MS. How to measure survey reliability and validity. SAGE Publications, London, 1995.

7. Carmines EG, Zeller RA. Reliability and Validity assessment. SAGE Publications, London, 1979.

8. Kmom. Prenatal Testing: Ultrasound Safety and Accuracy. http://www.plus-size-

pregnancy.org/Prenatal%20Testing/prenataltestultrasoundsafety.htm Retrieved 2010-02-12.

9. Olesen AW, Westergaard JG, Thomsen SG, Olsen J. Correlation between self-reported gestational age and ultrasound measurements. Acta Obstetricia et Gynecologica Scandinavica. 2004;83(11):1039-1043.

10. Kieler H, Axelsson O, Nilsson S, Waldenstrom U. Comparison of ultrasonic measurement of biparietal diameter and last menstrual period as a predictor of day of delivery in women with regular 28 day-cycles. Acta Obstetricia et Gynecologica Scandinavica. 1993;72(5):347-349.

11. Campbell S, Warsof SL, Little D, Cooper DJ. Routine ultrasound screening for the prediction of gestational age. Obetet Gynecol. 1985;65:613-620.

12. Persson PH, Weldner BM. Reliability of Ultrasound Fetometry in Estimating Gestational Age in the Second Trimester. Acta Obstetricia et Gynecologica Scandinavica. 1986;65(5):481-483.

13. O'Brien GD, Queenan JT. Growth of the ultrasound fetal femur length during normal pregnancy. Ammerical Journal of Obstetrics and Gynecology. 1981;141:833-837.

14. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on femur length at 10-25 weeks of gestation, and reference ranges for femur length to head circumference ratios. Acta Obstetricia et Gynecologica Scandinavica 2005;84:8:725-733.

USS Measurement	Period of Gestation (days)					
	Mean	SD	COV	F	Р	ICC (95% CI)
Biparietal diameter	134.6	11.5	8.5%			
Femur length	133.0	15.0	11.3%	6.5	0.001	0.7 (0.6 - 0.8)
Abdominal circumference	130.2	11.4	8.7%			
Head circumference	134.7	11.2	8.3%			

Table 1: Variation of period of gestation by Rater I

USS - ultra sound scan; SD - Standard deviation; COV - Coefficient of variation;

ICC - Intra-class correlation coefficient; CI - Confidence interval

Table 2: Variation of period of gestation by Rater II

USC Maagunamant	Period of Gestation (days)						
USS measurement	Mean	SD	COV	F	Р	ICC (95% CI)	
Biparietal diameter	131.7	13.6	10.3%				
Femur length	133.0	13.0	9.7%	4.80	0.003	0.8(0.7-0.9)	
Abdominal circumference	129.3	9.6	7.4%				
Head circumference	130.8	11.2	8.4%				

USS - ultra sound scan; SD - Standard deviation; COV - Coefficient of variation;

ICC – Intra-class correlation coefficient; CI – Confidence interval

Table 3: Variation of period of gestation by Rater III

USC Macaurament	Period of Gestation (days)					
USS Measurement	Mean	SD	COV	F	Р	ICC (95% CI)
Biparietal diameter	130.7	15.7	12.0%			
Femur length	127.5	17.1	13.4%	7.52	0.001	0.87
Abdominal circumference	127.6	12.0	9.4%			(0.8-0.91)
Head circumference	131.0	14.7	11.3%			

USS - ultra sound scan; SD - Standard deviation; COV - Coefficient of variation;

ICC – Intra-class correlation coefficient; CI – Confidence interval

Table 4: Distribution of the difference between the highest and lowest period of gestation by raters

Period of Gestation (Highest – Lowest)	Rater I n (%)	Rater II n (%)	Rater III n (%)
\leq 7 days	13 (22)	17 (28)	12 (20)
8 – 14 days	24 (40)	29 (48)	35 (58)
15 – 21 days	15 (25)	10 (17)	9 (15)
> 21 days	8 (13)	4 (7)	4 (7)
Total	60 (100)	60 (100)	60 (100)