Comparative Accuracy of Clinical Estimate versus Menstrual Gestational Age in Computerized Birth Certificates

Ghulam Mustafa, MD^a Richard J. David, MD^{a,b}

SYNOPSIS

Objective. This study compares gestational age data obtained by clinical estimate with data calculated from the date of the last menstrual period (LMP) as recorded on birth certificates.

Methods. The authors analyzed 476,034 computerized birth records from three overlap years, that is, those that contained both menstrual and clinical estimates of gestational age, concentrating on cases within the biologically plausible range of 20–44 weeks.

Results. The overall exact concordance between the two measurements was 46%. For +1 week it was 78%, and for +2 weeks it was 87%. Incidence of prematurity was 16% with menstrual gestational age, while it was 12% with clinical estimate. About 47% of the LMP-based preterm births were classified as term by clinical estimate. Eighty-three percent of clinical estimate–based preterms were also preterms by LMP-based gestation. Birthweight frequency distribution curves for LMP-based gestational age are bimodal, indicating probable miscoding of term births. An apparent over-representation of births coded as exactly 40 weeks by clinical estimate suggests rounding off near term for this method.

Conclusion. Agreement between menstrual and clinical estimates of gestational age occurs most often close to term, with significant disagreement in preterm and postterm births. Use of different methods of determining gestation in different years or geographic populations will result in artifactual differences in important indicators such as prematurity rate.

^aDivision of Neonatology, Cook County Hospital

^bDepartment of Pediatrics, University of Illinois at Chicago

Address correspondence to Richard J. David, MD, Division of Neonatology, Cook County Children's Hospital, 700 South Wood Street, Chicago, IL 60612

^{©2001} Association of Schools of Public Health

INTRODUCTION

Birthweight and gestational age have long been known to be the primary determinants of neonatal morbidity and mortality.¹ Conventionally, duration of pregnancy is calculated as the interval between onset of the last normal menses and the date of birth. This has been used as the definition of gestational age of the newborn,² although this definition overestimates the duration of pregnancy by approximately two weeks, the average interval from the beginning of the onset of the menstrual cycle to actual ovulation and conception.³ Nevertheless, the number of completed weeks from the onset of the last menstrual period (LMP) to birth has remained the standard definition, whether calculated by the obstetrician for a patient in the course of clinical management or computed by epidemiologists for members of a population.4-7

Beginning in 1939, the National Office of Vital Statistics published model certificates of live birth, which provided for inclusion of "number of months of pregnancy," an estimate by the clinical personnel involved. With the next revision of the standard certificates in 1949, weeks replaced months as the time measure of this clinical estimate.⁸ In the 1960s, birth certificates in many states began calling for the date of the onsent of the last normal menses in place of the number of weeks of pregnancy. Population-based studies comparing the distributions of gestational age values obtained by clinical estimate and by LMP had shown a clear preference for the latter. Anomalies of the distribution of gestational age obtained by clinical estimate included obvious digit preference (higher frequencies with even weeks) and as many as 87% of values concentrated in the single week's category for 40 weeks, a finding considered unrealistic.9-11

However, the limitations of LMP-based gestational age determinations have also been reported widely in the scientific literature. Several factors may lead to an error in gestational age derived from the LMP, including irregularities of menstrual cycle, individual variation in length of the cycle, preconception amenorrhea following oral contraceptives, implantation bleeding or other bleeding early in pregnancy, and recall errors by mothers.^{12,13} Kramer et al., based on 11,045 pregnancies with second trimester ultrasound data, pointed out that LMP alone systemically overestimates the incidence of preterm and postterm delivery.¹⁴ In addition, approximately 20% of live birth certificates in the United States have missing or incomplete data for LMP. Birth certificates for children born to women of low socioeconomic status or receiving late prenatal care contain a greater proportion of missing or implausible LMP dates.^{15,16} This is unfortunate, since a higher rate of adverse outcomes makes these women a particularly important group to study if population-wide improvements in perinatal health are to be achieved.

In spite of all these obstacles, researchers have used length of gestation from LMP to construct intrauterine growth curves and to calculate prematurity rates, inter-pregnancy intervals, and adequacy of prenatal care in populations.^{1,17-20} The LMP was also the original basis for developing various clinical measures of gestational age, including fetal ultrasonographic dating²¹⁻²³ and newborn infant gestational age assessments.²⁴⁻²⁶

Clinical estimation of gestational age was reintroduced in Illinois birth certificates in 1989. Beginning in 1992, presumably for reasons of confidentiality, the state dropped the exact dates of birth from the computerized birth certificates made available to researchers in this state. This practice makes it impossible to calculate gestational age from LMP. We therefore undertook a study of births from the years 1989 to 1991, the overlap period, to compare the two methods. The objective of our study was to investigate the concordance between these two measures and, where discrepancies were revealed, to determine as far as possible which of the two measures was most likely to be accurate.

METHODS

Population

In 1990, Illinois had a population of about 11.4 million, 70% of whom lived in urbanized areas. The ethnic composition was 75% non-Hispanic white, 15% black, 8% Latino, and 2% Asian. Live births reported in the state ranged from 187,026 to 192,349 during the years of study. More than 99% of births occurred in hospitals.

Overall completeness and plausibility analysis

We used computerized birth certificates for the years 1989 to 1991 obtained from the State of Illinois Department of Health, selecting only singleton live births for initial investigation (n = 568,488). We calculated gestational age in completed weeks by subtracting LMP from date of birth. We excluded women who did not have complete date of LMP. Information on cycle length, regularity, and contraception were not available. Hospital personnel recorded clinical estimate in completed weeks, and it appears on the birth records in that format. Although this estimate may have included ultrasound data as well as results of newborn

Characteristic	Menstrual estimate	Clinical estimate
Within range (20-44 wk)	477,763 (84.0%)	565,530 (99.5%)
Out of range (<20 or >44 wk)	11,660 (2.1%)	497 (0.1%)
Missing or incomplete	79,065 (13.9%)	2,461 (0.4%)
Total	568,488 (100.0%)	568,488 (100.0%)

Table 1. Comparison of missing and out of range data for single live births in Illinois, 1989-1991

examinations, birth certificates do not specify what information was used to make this determination.

Comparisons between methods

For final analysis, a subset of cases that contained both LMP-based and clinical estimations of gestational age within a range of 20 to 44 weeks were selected as biologically plausible for determination of concordance (n = 476,034). We first assessed plausibility of gestational measures within this range by a simple comparison of the percentages in preterm, term, and postterm categories. When the methods yielded different results, we included consideration of the birthweight, which was recorded on virtually 100% of certificates and is strongly related to gestational age.

Frequency distribution analysis

To explore the higher percentage of births classified as premature using the LMP method, we performed an analysis of birthweight frequency distributions. We created gestational age-specific birthweight distribu-



Figure 1. Distribution of menstrual and clinical estimates of gestational age

tion plots at various gestational ages, as determined by each of the two gestational age estimating methods. We then compared the resulting curves both for overall shape and for the characteristics of the two modes of bimodal distributions.

RESULTS

Overall completeness and plausibility

Table 1 shows the quality and completeness of data used for the analysis. Only 0.4% had missing clinical estimate, as compared to 14% that had missing or incomplete LMP. Table 2 and Figure 1 show the distribution characteristics of the two gestational age measures. Clinical estimate exhibits a slightly higher mean and one week higher median compared to menstrual estimate. This is related to the concentration of more than 41% of births in the single category of 40 weeks for the clinical estimate, where the LMP-based distribution is more rounded, with its peak spread over 39 to 40 weeks. The LMP distribution has a larger standard deviation and a wider range between the 5th and 95th percentiles of distribution, while the concentration of cases at the mode in the clinical estimate distribution is reflected in much higher kurtosis.

Table 2. Variation between menstrual and clinicalestimates of gestational age and its effects ongestational age-based categories

Characteristic	Menstrual estimate	Clinical estimate
Mean (weeks)	38.9	39.1
Median (weeks)	39	40
Standard Deviation	2.41	2.09
5th and 95th percentiles	35-42	36-41
Skewness	-2.56	-3.82
Kurtosis	11.47	17.53
Extremely preterm (<33 weeks)	2.2%	1.6%
Preterm (< 37 weeks)	9.8%	6.7%
Term (37-42 weeks)	86.7%	92.9%
Post-term (>42 weeks)	3.5%	0.4%



Comparisons between methods

The overall exact concordance between the two measurements, that is, having the same gestational age week value for both menstrual and clinical estimates, was 46%. Within +1 week it was 78%, and for +2 weeks it was 87%. However, the agreement between the two measures varied considerably based on gestational age. The highest exact concordance was observed at 40 weeks (72%). For more than 40% of cases in the range of 27 to 34 weeks, clinical estimate exceeded the menstrual gestation by more than two weeks.

We found a higher percentage of preterm births (gestation <37 weeks) and postterm (after 42 weeks) births using the menstrual estimate (Table 2). Only 55% of the LMP-based preterms were also classified as preterms by clinical estimate. Indeed, 61% of these supposedly preterm births (based on LMP) had birthweights >3,000 grams, a suspiciously high birthweight for truly premature infants. On the other hand, of the births coded as preterm by clinical estimate, 79% were also coded as preterm by LMP-based gestation estimate.

Frequency distribution analysis

To explore this discrepancy between gestational age estimates, we plotted distributions of birthweight for each gestational week for both techniques. A typical example is shown in Figure 2, a comparison between the two techniques for 33 weeks. Similar results were found at all gestational ages <36 weeks. As Figure 2 demonstrates, the LMP-based distribution exhibits a bimodal pattern with a long right tail, while the clinical estimate distribution curve is unimodal and symmetrical. The birthweight of the single mode of the clinical estimate curve was close to but consistently somewhat below the lower weight mode of the bimodal LMP-based curve for all gestational ages <36 weeks. The second (higher birthweight) mode of the menstrual estimate curve had no parallel in the clinical estimate curve, suggesting that these second modes may be an artifact.

In Figure 3 we plotted three bimodal LMP curves from different preterm gestation categories. The first mode in the three curves, which in general corresponds to the clinical estimate modes, increases from about 750 grams at 25 weeks to 1,500 grams at 30 weeks and to 2,750 grams at 35 weeks. The second mode, seen distinctly at around 3,250 grams for 25 weeks and 30 weeks also shows up as an upward skewness in the 35-week curve (Figure 3). The second mode was present at the same weight in the 33-week gestation curve (Figure 2). When we analyzed these second mode cases separately, only 6% of them had clinical gestational age estimates of prematurity comparable to that indicated by the LMP on the same record, again supporting the conclusion that they constitute a coding error of some type.

Birthweight distribution curves for gestational age >36 weeks were similar for both techniques of gestation estimate. However, as noted above, 41.2% of the term births (37–42 weeks) were coded as exactly 40 weeks using clinical estimate, versus 26.7% of records with LMP-based gestational ages.

DISCUSSION

Evaluation of gestational age in unselected populations is an important component of perinatal outcome assessment. However, in this study the menstrual and clinical estimate of gestational age—two measurements applied in the analysis of large populations exhibited notable discrepancies. It was not clear from our findings that either one of these techniques used alone would prove optimal over the complete range of gestational age values.

Alexander et al. have pointed out that discrepant results from these two methods of estimating gestational age could lead to discontinuities in tracking population outcome measures such as prematurity rates across time.7 Similarly, invalid comparisons would result when comparing data from two states using different methods. Our results support this prediction. The National Center for Health Statistics recommends using the clinical estimate as an adjunct to gestational age computed from LMP,27 and some researchers have adopted the approach of substituting clinical estimates only for those records containing no-or biologically implausible—LMP data.^{28,29} This approach would solve the problem of significant subgroups missing LMP information (14% in our data) with minimum problems of discontinuity across time and space. However, it requires that data for both methods of estimating gestational age be present on the birth records available to researchers, a condition no longer met in Illinois after the years we studied. Most of the births missing LMP data are missing only the day, so that another approach for most incomplete records would be imputation using a standard algorithm.³⁰ The relative strengths of this approach compared to clinical estimate require further study.

The concordance between LMP-based and clinical estimates varies over the gestational range. The maximum concordance is at or near term. The discrepancy becomes quite marked as the gestational age deviates



Figure 3. Distribution of birth weights at 25, 30, and 35 weeks gestation based on LMP. The arrow points to the second mode, which stays at about the same point in spite of increase in gestational age.

progressively towards earlier or later dates. The overall concordance appears to be high because the vast majority of babies are born at or near term, as previously pointed out by Kramer et al.¹⁴ However, it is precisely in infants who are born pre- or postterm that accurate gestational age estimation is of greatest significance for health outcomes.

Of particular interest are the conflicting results for preterm births. The bimodal birthweight distributions we demonstrated using menstrual gestation estimates are similar to those previously reported in other populations.^{16,31,32} We have shown previously that miscoding the LMP of a random 2% of all term births would account for the records found in all the second modes seen in such bimodal curves as menstrual gestational ages below term.¹⁶ The current study confirms, at least as far as the clinicians involved are concerned, that in more than 90% of cases these second mode births are not as premature as the LMP would indicate. Various data cleaning techniques have been devised to eliminate the presumably miscoded second mode cases for purposes of generating intrauterine growth curves,^{17,33,34} but this study raises questions about the first mode cases as well. These first ("true") birthweight modes were still heavier than the corresponding clinical estimate modes (Figure 2). Either of these sets of distributions-"cleaned" LMP-based curves (essentially, curves like those in Figure 3 with the second modes removed), or clinical estimate-based curves-could be used to produce realistic intrauterine growth curves, but they would not agree. Which represents the true standard for the population?

Unfortunately our data cannot adequately address the question of which gestational age measure most faithfully represents the true distribution for the population because there is no agreed upon "gold standard" for pregnancy dating. Although clinical estimate has few biologically implausible out-of-range values as assessed by birthweight, one could argue that knowledge of birthweight after delivery would naturally influence the clinical estimation recorded on the birth certificate, whether or not this was justified in every case. This could result in erroneously editing out the true variability present in the population. There has been a considerable improvement in the plausibility of clinical estimates from their earlier appearance on certificates in the 1950s, when more than 80% of all births were reported at exactly 40 weeks. This may reflect the widespread use of prenatal ultrasound today. Nevertheless, some tendency to clump at 40 weeks persists in the current data. The accuracy of clinical estimates beyond their general pattern of conforming to prior expectations remains unknown.

Many view early prenatal ultrasound obtained at around 10 weeks as the new "gold standard" because there is little variation in the size of the embryo at this stage of development.35 However, the lingering conceptual problem with scanning is that what is being measured is size, not time, and after the earliest stage of pregnancy individual growth differences become evident. For example, Moore et al. have shown in a longitudinal analysis that a fetus whose bi-parietal diameter is one standard deviation above the mean at 16 weeks is similar to a fetus whose corresponding measurement is one standard deviation below the mean at 18 weeks.³⁶ Sex differences in fetal size have been claimed as early as 8 and 14 weeks.^{36,37} Although an early scan is clearly useful when dates are uncertain or when the clinical findings are not in keeping with the menstrual data, scanning at 8 to 10 weeks is by no means routine. Although it is estimated that 70% of pregnant women in the United States received an ultrasound in 1996,³⁸ most of these studies were too late in gestation for precise dating.36,37 The study by Kramer and co-workers comparing ultrasound dates with LMP, the main attempt so far to apply sonographic dates on a population basis, used ultrasound studies obtained at an average of 16-18 weeks.14

It appears from our analysis that one technique for determining gestational age in populations will not be optimal for all requirements. Babies of premature gestation are probably most reliably identified using the clinical estimate, while for babies at or near term the LMP-based gestational age determinations produce a more realistic distribution. However, the dating errors near term are probably minor and could safely be ignored for many purposes. Whether clinical estimate underestimates or LMP-derived gestational age overestimates the number of post-term infants remains an open question. A more definitive evaluation of gestation-estimating techniques for large unselected populations will require application of a different technique, such as very early ultrasound, to every member of a large population. Such a study would require a different pattern of access to prenatal care than that currently achieved in this country.

REFERENCES

- Lubchenco LO, Hansman C, Dressler M, Boyd E. Intrauterine growth as estimates from live born birthweight data at 24 to 42 weeks gestation. Pediatrics 1963;32:793-800.
- Silverman WA. Nomenclature for duration of gestation, birthweight and intrauterine growth. Pediatrics 1967;39: 935-9.

- 3. Treloar AE, Behn BG, Cowan DW. Analysis of the gestational interval. Am J Obstet Gynecol 1967;99:34-45.
- 4. Chiswick ML. Commentary on current World Health Organization definitions used in perinatal statistics. Br J Obstet Gynaecol 1986;93:1236-8.
- 5. Hall MH. Definitions used in relation to gestational age. Paediatr Perinat Epidemio. 1990;4:123-8.
- 6. Emery ES 3rd, Eaton A, Grether JK, Nelson KB. Assessment of gestational age using birth certificate data compared with medical record data. Paediatr Perinat Epidemiol 1997;11:313-21.
- Alexander G, Tompkins ME, Petersen DJ, Hulsey TC, Mor J. Discordance between LMP-based and clinically estimated gestational age: Implications for research, programs, and policy. Public Health Rep 1995;110:395-402.
- 8. Vital Statistics of the United States. 1950;1:101-3. 1953;2:270-300.
- 9. Matthew AM. Birthweight and length of gestation with relation to prematurity. JAMA 1951. 146:897-901.
- Schwartz S, West H. Potentialities and limitations of medical data on official birth certificates. AJPH 1960; 50:338-45.
- Hammes LM, Treloar AE. Gestational interval from vital records. AJPH 1970;60:1496-1505.
- 12. Fraser IS, Wiesberg E. Fertility following discontinuation of different methods of fertility control. Contraception 1982;26:389-415.
- Gjessing HK, Skjoerven R, Wilcox AJ. Errors in gestational age: Evidence of bleeding early in pregnancy. AJPH 1999;89:213-8.
- 14. Kramer MS, McLean FH, Boyd ME, Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm and postterm gestations. JAMA 1988;260:3306-8.
- Buekens P, Delvoye P, Wollast E, Crobyn. Epidemiology of pregnancies with unknown last menstrual period. J Epidemiol Community Health 1984;80:38-9.
- 16. David RJ. The quality and completeness of birthweight and gestational data in computerized birth files. AJPH 1980;70:964-73.
- 17. David RJ. Population-based intrauterine growth curves from computerized birth certificates. South Med J 1983;76:1401-6.
- Arbuckle TE, Wilkins R, Sherman GJ. Birthweight percentiles by gestational age in Canada. Obstet Gynecol 1993;8:39-48.
- Alexander GR, Cornley DA. Prenatal care utilization: its measurement and reationship with pregnancy outcome. Am J Prev Med 1987;3:243-53.
- 20. Khoshnood B, Lee KS, Wall S, Hsieh HL, Mittendorf R. Short pregnancy intervals and the risk of adverse birth outcomes among five racial/ethnic groups in the United States. Am J Epidemiol 1998;148:98-805.
- Altman DG, Chitty LS. Charts of fetal size: 1. Methodology. Br J Obstet Gynaecol 1994;101:29-34.

- 22. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. Br J Obstet Gynaecol 1994;101:35-43.
- Person PH, Grennert L, Genneser G, Gullberg B. Normal range curves for the intrauterine growth of the biparietal diameter. Acta Obstet Gynaecol Scand Suppl 1978:15-20.
- 24. Dubowitz LMS, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. J Pediatr 1970;77:1-10.
- Ballard JL, Novak KK, Driver M. A simplified score for assessment of fetal maturation of newly born infants. J Pediatr 1979;95:769-74.
- Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard Score, expanded to include extremely premature infants. J Pediatr 1991;119: 417-23.
- Hospitals' and physicians' handbook on birth registration and fetal death reporting. National Center for Health Statistics, Hyattsville, MD, 1987.
- Zhang J, Bowes WA. Birthweight-for-gestational-age patterns by race, sex, and parity in the United States population. Obstet Gynecol 1995;86:200-8.
- 29. Kiely JL. What is the population-based risk of preterm birth among twins and other multiples? Clinical Obstet Gynecol 1998;41:3-11.
- Taffel S, Johnson D, Heuser R. A method of imputing length of gestation on birth certificates. Vital Health Stat 2, No. 93. DHHS Pub. No. (PHS) 82-1367. Public Health Service. Washington. U.S. Government Printing Office, May 1982.
- 31. Grunewald P. Growth of the human fetus: 1. Normal growth and its variation. Am J Obstet Gynecol 1966;94: 1112-9.
- Milner RDG, Richards B. An analysis of birthweight by gestational age of infants born in England and Wales, 1967-1971. J Obstet Gynaec Birth Cwlth 1974;81:956-66.
- Williams RL. Intrauterine growth curves: Intra- and international comparisons with different ethnic group in California. Prev Med 1975;4:163-72.
- Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. Obstet Gynecol 1996;87:163-68.
- Cunnigham FG, MacDonald PC, Leveno KJ, Gant NF, Gilstrap III LC. Williams Obstetrics. 19th ed. Norwalk (CT): Appleton & Lange; 1993.
- Moore WMO, Ward BS, Jones VP, et al. Sex differences in fetal head growth. Br J Obstet Gynecol 1988;95:238-42.
- 37. Pedersen JF. Ultrasound evidence of sexual differences in fetal size in first trimester. BMJ 1980;281:675-77.
- American College of Obstetricians and Gynecologists news release; accessed from URL: http://www.acog .org.com/ April 15, 1999.