1990

Gestational Age Reporting and Preterm Delivery

Greg R. Alexander

MARK E. TOMPKINS
University of South Carolina, tompkins.mark@sc.edu

Donald A. Cornely

Publication Info
Published in *Public Health Reports*, Volume 105, Issue 3, 1990, pages 267-275.

Natality Survey. Fam Plann Perspect 17: 118–124, May/June 1985.
16. Fingerhut, L. A., Makuc, D., and Kleinman, J. C.: Delayed prenatal care and place of first visit: differences by health insurance and education. Fam Plann Perspect 19: 212–234, September/October 1987.
17. Rosner, B.: Fundamentals of biostatistics. Ed. 2. Duxbury Press, Boston, 1986.
18. Cooney, J. P.: What determines the start of prenatal care? Prenatal care insurance and education. Med Care 23: 986–997, August 1985.

Gestational Age Reporting and Preterm Delivery

GREG R. ALEXANDER, MPH, ScD
MARK E. TOMPKINS, PhD
DONALD A. CORNELY, MD, MPH

Dr. Alexander is an Assistant Professor in the Department of Maternal and Child Health, School of Hygiene and Public Health at The Johns Hopkins University and a Clinical Assistant Professor in the Department of Preventive Medicine and Community Health, School of Medicine at the University of South Carolina. He is also the Director of the Region III Perinatal Information Consortium. Dr. Tompkins is an Associate Professor in the Department of Government and International Studies and an Adjunct Associate Professor in the Department of Preventive Medicine and Community Health, School of Medicine at the University of South Carolina. Dr. Cornely is a Professor in the Department of Maternal and Child Health, School of Hygiene and Public Health at The Johns Hopkins University.

Tearsheet requests to Greg R. Alexander, MPH, ScD, Maternal and Child Health, School of Hygiene & Public Health, The Johns Hopkins University, 624 North Broadway, Baltimore, MD 21205.

The computation procedure used to impute gestational ages by the 'Preceding Case' method is available from the authors upon request. The procedure is written in SAS statistical computing language.

Synopsis

This study examines recent trends in the reporting completeness and quality of gestational age estimates derived from the date of the last normal menses (DLNM) as reported in South Carolina vital records from 1974 to 1985. Noteworthy improvements in the completeness of reporting emerged during this period with a decline from 31.1 percent missing information in 1974 to 6.6 percent missing in 1985. Completeness of reporting and strategies for imputing values for missing data were analyzed for their impact on the calculation of the percentage of preterm live births. The results indicate that the underreporting of gestational age can lead to marked underestimation of the preterm percentage in a population and to misinterpretation of trends in these percentages.

Based on the results of this analysis, it is recommended that preterm percentages be based on cases with DLNM gestational age values between 20 and 50 weeks. Since cases with missing or implausible gestational age data have a greater risk of a poor pregnancy outcome, these findings emphasize the importance of identifying both the completeness of data reporting and the use of imputation and deletion strategies when employing population-based DLNM data to calculate gestational age related indicators.

ACCURATE DETERMINATION OF gestational age is important in obstetric and pediatric clinical practice (*1, 2*). In public health policy and research, accurate population based gestational age data are needed for a variety of purposes. These include monitoring the incidence of preterm delivery and intrauterine growth retardation, investigating the potential risk factors associated with preterm birth, constructing prenatal care use indices, and evaluating interventions focused on the prevention of preterm labor and delivery. Unfortunately, the precise determination of the duration of pregnancy in the human female represents a formidable task.

The clinical use of the date of last normal menses (DLNM) to estimate the duration of gestation has been established for well over 130 years (*3*). The gestational age interval, as calculated from the DLNM and the date of delivery, has been used to assess the validity of a number of antenatal and postnatal gestational age estimation procedures (*4–8*). Although the interval based on the DLNM is considered to exceed the interval from conception to delivery by approximately 2 weeks (*9*), the gestational interval has become the standard measurement to describe duration of pregnancy and the gestational age of the infant.

Although the use of the DLNM to calculate the gesta-

Table 1. Percent distribution by gestational age of single live births based on date of last normal menses to South Carolina residents by year, 1974–85

Year	Missing	Gestational age interval (weeks)					Number
		<20	20–36	37–44	45–50	>50	
1974	31.09	0.18	7.30	59.45	1.73	0.24	47,525
1975	17.86	0.27	9.01	70.20	2.29	0.38	45,769
1976	14.99	0.26	10.03	71.82	2.36	0.54	46,709
1977	10.90	0.19	10.34	75.45	2.63	0.48	48,789
1978	3.60	0.99	10.98	81.13	2.71	0.60	48,555
1979	5.03	0.17	10.87	80.60	2.73	0.60	49,507
1980	5.51	0.20	10.92	80.54	2.32	0.51	50,878
1981	3.88	0.20	11.28	81.79	2.34	0.50	50,665
1982	4.35	0.22	11.11	81.42	2.37	0.54	50,523
1983	4.42	0.20	11.09	81.64	2.29	0.35	49,674
1984	5.54	0.17	10.69	80.92	2.38	0.32	49,496
1985	6.58	0.17	10.95	79.71	2.28	0.31	50,691

Table 2. Biennial percentages of live births to South Carolina residents with unusable[1] data on gestational age interval, 1975–85

Category	1974–75	1976–77	1978–79	1980–81	1982–83	1984–85
Total	25.13	13.64	5.50	5.41	5.04	6.55
Maternal race and marital status:						
White-wed	19.66	11.18	4.24	4.06	4.12	5.06
White-unwed	28.37	17.48	9.34	8.43	8.29	10.29
Nonwhite-wed	30.66	15.15	5.98	5.92	4.73	6.51
Nonwhite-unwed	36.25	19.00	7.98	8.21	7.28	9.84
Missing	47.37	36.00	8.06	6.60	5.87	7.92
Maternal age (years):						
10–17	32.11	18.56	8.78	8.07	7.74	11.03
18–34	24.24	13.01	5.19	5.17	4.79	6.22
Older than 34	25.05	14.96	4.93	5.17	5.48	6.01
Missing	55.56	45.45	50.00	70.00	71.43	50.00
Maternal education:[2]						
0–8 grades	30.15	17.62	7.04	7.18	7.83	8.69
9–11	27.57	15.27	6.43	6.58	6.39	8.09
12	23.06	12.35	4.95	4.99	4.59	6.28
More than 13	20.43	10.97	3.94	3.97	3.76	4.82
Missing	42.83	27.47	22.06	33.93	27.27	27.72
Hospital size:						
Less than 500 beds	29.92	19.14	4.97	4.22	4.42	5.89
500–999	16.97	15.23	6.60	7.01	5.49	7.74
1,000–1,999	29.42	12.66	3.70	3.06	4.51	6.95
More than 2,000 beds	17.86	9.12	5.19	5.65	3.63	4.40
Out-of-State	17.10	13.91	15.81	14.77	16.10	16.71
Missing	29.55	18.22	9.32	9.57	11.85	16.74

[1] A birth is considered to have unusable gestational age interval data if the date of last normal menses is missing or if the calculated gestational age interval is less than 20 or greater than 50 completed weeks.
[2] Excludes women under 18 years of age.

tional interval has a long history, problems with reporting completeness and quality have persisted. Digit preference (multiples of 5) in reporting DLNM has been observed (*10, 11*). It has been noted that the calculated DLNM gestational interval can result in a range of values that extends beyond what is generally viewed as biologically plausible and in a gestational age value that is inconsistent with an infant's birth weight (*12–20*). A number of reports using vital records indicate 2 percent or more of infants with a gestational age beyond 45 weeks (*18, 20, 21*). In addition to errors in recording the DLNM, possible explanations for extraordinary gestational age interval values include variations in the preovulatory interval and misidentification of the actual DLNM by the female due to sporadic bleeding, previous unrecognized abortions, or other factors (*9, 15–18, 22*). Despite these inherent limitations on data quality, the use of DLNM has been endorsed by researchers for its ability to provide a useful gestational age estimate (*6, 7, 9, 23*) and, for vital record purposes, has been preferred over physician estimates of gestational age (*24–27*).

On both vital and medical records, DLNM information has commonly been missing, involving both unreported and incompletely reported DLNM (*12, 14, 18, 21, 28–30*). In the United States during the 1970s, the DLNM was completely reported on approximately 80 percent of vital record reports (*21*). U.S. vital record data from the late 1960s indicated regional variation in completeness of reporting, with missing data being a more common problem in the South (*31*). A number of researchers have further noted that reporting completeness of DLNM varies across population subgroups (*18, 28–30*). Women with missing DLNM data tend to be at lower socioeconomic status and at higher medical risk of poor pregnancy outcomes than those with complete DLNM data (*18, 21, 28, 29, 32*). These reports suggest that the exclusion of persons with missing DLNM from calculations of preterm or small-for-gestational-age percentages may result in an appreciable underestimation of the true extent of these conditions in a population.

A number of strategies have been suggested for addressing missing and presumed inaccurate data on DLNM (*19, 21, 30, 33*). Since a large proportion of records with missing DLNM information lack only the day of last normal menses (LNM) and have the remainder of the LNM date recorded (*21*), some researchers have imputed a value (usually 15) for the missing day (*34, 35*). This technique fails to address the problem of implausible or inconsistent gestational age values, and it has been noted to introduce bias into the gestational age distribution (*21*). Approaches that establish gestational age as a function of birth weight have also been employed (*19, 21, 33*), but it is uncertain how this strategy may influence comparisons of birth weight-gestational age specific mortality or the distribution of birth weight among gestational age intervals when studying intrauterine growth.

In this study we examine recent trends in the reporting completeness and quality of gestational age estimates derived from DLNM as contained in South Carolina's live birth records for the period 1974–85. Temporal variations in reporting completeness are assessed by specific population characteristics including maternal age, race, education, and marital status, as well as by hospital delivery size categories. The effect of completeness of reporting and the impact of strategies to impute missing data and to discard improbable data are then compared in calculations of the percentage of preterm live births.

Methods

Data for this investigation were obtained from South Carolina live birth computer files for the years 1974–85. These computer files were made available for public access by the South Carolina Office of Vital Records and Public Health Statistics. Single live births to South Carolina residents were selected for examination.

Gestational age was calculated as the interval from the beginning DLNM to the date of birth. Following recommended convention (*36*), this gestational interval was truncated into completed weeks for analytical purposes. In specific instances, a value of 15 was imputed for the day of the beginning of last normal menses for those births for which this single missing piece of information prevented the calculation of the gestational age interval. Alternatively (*21*), a missing gestational age was imputed by inserting the gestational age of a preceding birth with a corresponding month of LNM, a 500-gram birth weight interval, and race of mother. Preterm percentages based on the 'Day 15' imputation method, the 'Preceding Case' method, or no imputed data are noted.

Preterm percentages are calculated as the proportion of total single live births occurring at less than 37 weeks. However, the actual denominator (live births with gestational age interval data available) for the percentages will vary according to the stated strategies employed to address missing and implausible gestational age data.

Results

Annual percentage distributions of gestational age for the years 1974–85 are reported in table 1. The gestational age of single live births to South Carolina resident mothers was estimated from the DLNM interval only for births with a complete DLNM. In 1974, a gestational age interval could not be determined for 31.09 percent of single live births due to missing DLNM data. A gestational interval of less than 20 completed weeks was calculated for 0.18 percent of the live births, and an interval in excess of 50 weeks was derived for 0.24 percent of the births.

In subsequent years, DLNM reporting improved markedly. Declining rapidly from 31.09 percent missing in 1974, the proportion of live birth records with missing DLNM was 3.6 percent by 1978. The annual proportion of records with complete DLNM information remained above 93 percent through 1985.

Temporal variations in gestational age reporting were examined by demographic and hospital delivery size groups. Table 2 provides biennial percentages of live births with unusable gestational age information defined as either missing the DLNM or having a calculated gestational age considered implausible. A rather generous criteria for implausible gestational age values (less than 20 weeks or more than 50 weeks) was employed. In

Table 3. Characteristics of gestational age reporting for live births to South Carolina residents, 1974-85

Characteristics	1974–77		1978–81		1982–85	
	Number	Percent	Number	Percent	Number	Percent
Total live births	188,792	...	199,605	...	200,384	...
No imputation:						
Missing	35,271	18.68	9,011	4.51	10,472	5.23
Implausible[1]	1,198	0.63	1,873	0.94	1,135	0.57
Unusable[2]	36,469	19.32	10,884	5.45	11,607	5.79
Day 15 method:						
Unusable[2]	4,386	2.32	4,206	2.11	4,349	2.17
Efficiency[3]	...	87.56	...	53.32	...	58.43
Preceding case method:						
Unusable[2]	4,460	2.36	4,198	2.10	4,344	2.17
Efficiency[3]	...	87.36	...	53.41	...	58.52

[1] Implausible gestational ages are defined as less than 20 and more than 50 weeks.
[2] Unusable gestational age values include both missing (or not imputed) and implausible.
[3] Efficiency is defined as the proportion of the original missing data (no imputation) recaptured as usable (20–50 weeks).

general within all subgroups, a substantial decline in the proportion of births with unusable gestational ages was evident between 1974–75 and 1982–83. However, a slight reversal in this apparent trend was observed in the last period, 1984–85. Variation in the proportion of unusable gestational age data was apparent among the subgroups. These proportions were highest among infants of unwed young women under 18 years and nonwhite or poorly educated (less than 12 years of completed education) mothers. A consistent reduction in the proportion of unusable gestational age data from live births to South Carolina resident mothers who were delivered out of State was not apparent throughout the 12-year period.

Temporal variations in digit preference in the reporting of day of LNM were examined. An LNM day ending in 0 or 5 was recorded for 30.7 percent of births in the 1974–77 period, for 31.1 percent in 1978–81, and for 29.7 percent in 1982–85. The random occurrence of these days is approximately 20 percent. While digit preference for these days was clearly evident, a temporal increase in this practice was not apparent as completeness of reporting improved.

Table 3 provides information on the proportion of total live births with missing DLNM data and the proportion of births for which a gestational age could be imputed. For those with missing DLNM information, a gestational age was derived by the 'Day 15' and the 'Preceding Case' method. During the 1974–77 period, 18.7 percent of single live births had a missing DLNM and 0.63 percent had an implausible gestational age. The application of the 'Day 15' method resulted in the calculation of a usable gestational age value between 20 and 50 weeks for 87.6 percent of the 35,271 cases with a missing DLNM. The 'Preceding Case' method recaptured as usable 87.4 percent of the missing cases. As reporting completeness improved in subsequent time periods, the percentage of cases with missing DLNM data that could be recaptured for use by either imputation method declined.

In figure 1, the impact of several strategies for handling missing and implausible gestational age data is displayed. Annual percentages of preterm births (less than 37 weeks) were calculated using three separate treatments of the gestational age interval data: (*a*) gestational ages from births with complete DLNM data and no imputation for missing data, (*b*) gestational ages for births with complete or 'Day 15' imputed DLNM data, and (*c*) gestational ages from births with complete or 'Preceding Case' imputed data. For each of these groups, preterm percentages were also calculated using only cases with gestational age intervals within the 20- to 50-week range.

In general, the exclusion of births with values less than 20 and greater than 50 weeks from these calculations resulted in a slight decrease in the annual preterm percentages. A more pronounced impact of employing only plausible data was observed in 1978, the year with the substantially higher proportion of less than 20-week gestational age live births (table 1). Using only data in the plausible range, the 'Day 15' method produced the highest annual preterm percentages, and the no imputation approach produced the lowest percentages.

The percentage of birth weight distributions for gestational age intervals was examined for two 4-year periods (1974–77 and 1982–85) when there were notably different levels of completeness of reporting the DLNM (table 4). This assessment was undertaken to determine if the observed increase in completeness of reporting was accompanied by an increased proportion of cases with inconsistent gestational age and birth weight values. Gestational age intervals were determined first by using no imputation and then by using the 'Day 15' method. Since the 'Preceding Case' method incorporates birth weight into its assignment of gestational age values, it was not used in this comparison.

For the 1974–77 period, 18.7 percent of South Carolina single live births had missing DLNM information. More than 10 percent of live births in the missing gestational age category were low birth weight (less than 2,500 grams), compared with 8 percent low birth weight for all live births. Further, more than 10 percent of the births with a gestational age interval less than 13 weeks and greater than 50 weeks were low birth weight. The highest proportion of live births weighing 4,251 grams or more was found in the 45- to 50-week group. The highest percentages (57.1 for less than 2,500 grams and 70.6 for less than 1,500 grams birth weight) were found in the 20- to 24-week gestational age interval group.

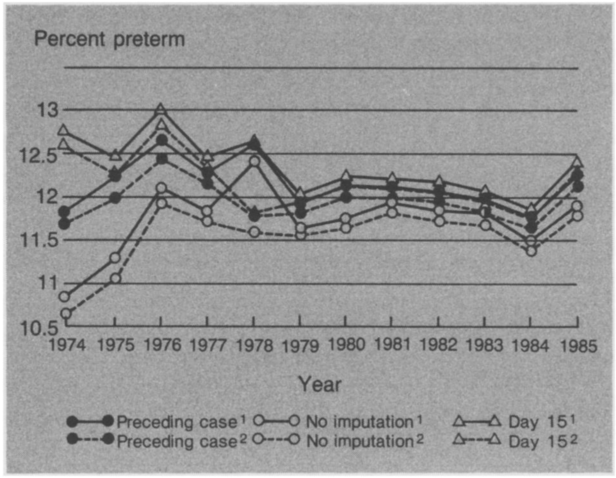
After including births with 'Day 15' imputed DLNM data, only 1.4 percent of live births in the 1974–77 period had missing gestational age data. For these remaining births with missing gestational age data, more than 15 percent were low birth weight. An appreciable change in the birth weight percentage distributions within each gestational age interval was not evident when births with imputed gestational age data were included with births with complete DLNM. Inconsistent birth weight and gestational age combinations, for example, more than 2,500 grams and 13–24 weeks, was apparent for both groups in roughly similar proportions, although an increase in the proportion of births with a gestational age interval of more than 50 weeks was observed for the 'Day 15' method group. A temporal increase in inconsistent values was not apparent.

Discussion

Noteworthy improvements in the completeness of reporting of DLNM were evident over the 12 years of data that we investigated. The quality of gestational age data did not appear to deteriorate with more complete reporting. Given a national percentage of reporting completeness of DLNM in 1983 of approximately 80 percent (*37*), these trends in South Carolina indicate that much national improvement is possible. Because reporting involves vital records personnel, hospitals staffs, and practicing clinicians, these data have implications for all groups.

Although the specific factors underlying the evident improvement in completeness of gestational age reporting cannot be precisely identified, several changes in vital record reporting procedures were implemented during this period. During the 1970s, the South Carolina Office of Vital Records and Public Health Statis-

Annual preterm percentages by method of imputing gestational age



[1] All data. [2] 20-50 weeks.

tics implemented more intensive centralization of quality control efforts, focusing on reducing invalid, missing, and in particular, unknown data entries. Follow-back inquiries to local personnel about initially reported unknown DLNM values were a major part of these activities. Further, ongoing statewide research and intervention projects, aimed at reducing low birth weight and preterm delivery, may have indirectly contributed to these trends by drawing the attention of hospital staff and clinicians to the issues of data quality and completeness.

As more attention is given to the prevention of preterm delivery, a standard criterion for the calculation and monitoring of the incidence of this problem in populations is needed. We suggest tentatively that preterm percentages for large populations be based only on births with a gestational age interval value between 20 and 50 weeks, if DLNM is used for the determination of gestational age. However, the use of this broad criteria (20 to 50 weeks) is a heuristic device that can be justified only partially.

In more than half of the live births with a gestational age of less than 20 weeks, a birth weight of 2,500 grams or greater was reported. Since many of these cases probably involved invalid gestational age data and may not be preterm births, their inclusion in the numerator of preterm percentage calculations is undesirable, and it is likely to result in an artificial inflation in preterm percentages. These gestational age categories typically comprised less than 1 percent of the data, but a sudden increase in this proportion, as observed in 1978, can result in a pronounced increase in preterm percentages. The value of employing a complex formula to delineate "valid" data, which might take into account the consistency of an infant's gestational age, birth weight, and survival characteristics, must be measured

Table 4. Percent birth weight distributions by gestational age interval and DLNM imputation strategy for live births to South Carolina residents, 1974–77 and 1982–85

		Gestational age intervals in weeks							
Weight (grams)	*Missing*	*<13*	*13–19*	*20–24*	*25–36*	*37–44*	*45–50*	*>50*	*Total*
Birth weight percentages for cases with complete DLNM data, 1974–77									
1–499	0.13	0.71	10.42	14.69	0.08	0.01	0.00	0.26	0.08
500–1,499	1.64	3.21	25.69	42.45	7.15	0.06	0.07	3.36	1.14
1500–2499	8.74	9.29	15.28	13.48	27.43	3.67	3.08	7.24	6.78
2500–4250	86.10	81.79	45.83	26.76	63.90	91.78	90.86	84.75	87.96
over 4251	3.15	5.00	1.39	1.21	1.11	4.44	5.90	4.26	3.92
Missing	0.23	0.00	1.39	1.41	0.33	0.05	0.09	0.13	0.12
Gestational age interval percentage	18.68	0.15	0.08	0.26	8.91	69.25	2.25	0.41	
Birth weight percentages for cases with complete 'day 15' imputed DLNM data, 1974–77									
1–499	0.37	0.63	11.32	12.98	0.10	0.01	0.00	0.35	0.08
500–1499	3.77	2.81	28.77	42.77	6.84	0.07	0.09	2.96	1.14
1500–2499	11.86	10.63	16.04	13.86	26.36	3.95	3.34	8.19	6.78
2500–4250	79.86	81.56	41.98	27.43	65.26	91.63	90.75	83.71	87.96
over 4251	2.85	4.38	0.94	1.33	1.11	4.29	5.72	4.70	3.92
Missing	1.29	0.00	0.94	1.62	0.34	0.06	0.11	0.09	0.12
Gestational age interval percentage	1.43	0.17	0.11	0.36	11.83	82.53	2.95	0.61	
Birth weight percentages for cases with complete DLNM data, 1982–85									
1–499	0.42	7.21	24.91	20.89	0.17	0.00	0.00	0.13	0.16
500–1499	2.71	6.31	22.26	46.11	8.00	0.07	0.13	1.58	1.27
1500–2499	9.32	24.32	9.43	8.65	27.87	3.28	3.23	4.61	6.23
2500–4250	84.21	59.46	41.51	24.23	62.94	91.71	90.77	87.09	87.89
over 4251	3.29	2.70	1.89	0.12	1.01	4.94	5.87	6.59	4.44
Missing	0.05	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.12
Gestational age interval percentage	5.23	0.06	0.13	0.40	10.56	80.92	2.33	0.38	
Birth weight percentages for cases with complete or 'day 15' imputed DLNM data, 1982–85									
1–499	0.87	6.02	24.66	20.46	0.17	0.00	0.00	0.21	
500–1499	4.70	8.27	21.92	44.86	7.93	0.08	0.12	1.48	1.27
2500–2499	12.84	23.31	9.25	8.67	27.32	3.33	3.24	4.75	6.23
2500–4250	79.33	60.15	42.47	25.90	63.51	91.67	90.97	87.66	87.89
over 4251	2.18	2.26	1.71	0.12	1.05	4.92	5.67	5.91	4.44
Missing	0.07	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
Gestational age interval percentage	1.49	0.07	0.15	0.43	11.31	83.55	2.54	0.47	0.47

[1] Percentages in the total category are based on all births with birth weight data.

NOTE: DLNM = day of last normal menses.

against considerations of parsimony in the establishment of a practical criteria. With these issues in mind, the lower bound of 20 weeks is proposed.

One limitation of proposing explicit criteria to serve as a convention for calculating a health status indicator is the possibility that future advances will render the criteria obsolete. The proposal of 20 weeks as the lower limit of usable gestational age data for live births may become outdated with future advances in medical technology that can increase the survivability of infants with extremely preterm deliveries. A more fluid approach to this problem would be to link the criteria of a plausible gestational age value to the gestational age criteria for reporting a fetal death. The majority of States currently incorporate 20 weeks gestational age into their fetal death reporting regulations. As these criteria are lowered in the future, a reciprocal change in the convention for the calculation of the percentage of preterm births should be considered.

Cases with a gestational age interval between 45 and 50 weeks represent more than 2 percent of the data and, based on the numbers alone, their removal from preterm calculations would be troubling. As noted previously, this small but relatively persistent percentage of gestational age values of 45 to 50 weeks based on DLNM has repeatedly been identified (*18, 20, 21*), and improbably long gestational intervals have been suggested to represent errors in both recognizing and recording the DLNM (*12*). Although these gestational age interval values may be clinically implausible, the birth weight characteristics of this group are not untypical of term or post term births. As such, in this report they

Table 5. Biennial percentage of preterm live births to South Carolina residents by race of mother and data imputation strategy, 1974–85

Year	White				Nonwhite			
	No imputation	*Preceding case I*[1]	*Preceding case II*[1]	*Day 15*	*No imputation*	*Preceding case I*[1]	*Preceding case II*[1]	*Day 15*
				All gestational ages[2]				
1974–74	7.55	7.99	7.96	8.37	17.60	18.39	18.32	19.23
76–77	7.63	7.90	7.85	8.08	18.85	19.34	19.29	19.70
78–79	7.88	8.02	8.00	8.06	18.17	18.44	18.41	18.59
80–81	7.59	7.74	7.72	7.82	18.26	18.58	18.55	18.71
82–83	7.85	7.96	7.95	8.05	17.93	18.21	18.21	18.34
84–85	7.79	7.93	7.92	8.06	18.04	18.50	18.48	18.65
				Gestational ages 20–50 weeks[3]				
1974–75	7.37	7.83	7.81	8.23	17.34	18.15	18.14	18.98
76–77	7.50	7.75	7.74	7.95	18.66	19.14	19.13	19.52
78–79	7.48	7.61	7.60	7.66	17.65	17.91	17.91	18.07
80–81	7.53	7.67	7.67	7.76	18.05	18.34	18.36	18.49
82–83	7.81	7.91	7.91	8.01	17.66	17.95	17.95	18.07
84–85	7.74	7.88	7.87	8.00	17.82	18.26	18.27	18.41

[1] In addition to the selection criteria of a corresponding month of last normal menses (LNM), birth weight interval and race used in Preceding Case method I, Preceding Case method II further restricts the selection of a preceding case to those with gestational age values of 20 to 50 weeks.

[2] Preterm percentages based on the number of cases of less than 37 weeks divided by the total number of cases with available gestational age interval data.

[3] Preterm percentages based on the number of cases with a gestational age of 20 to 36 weeks divided by the number of cases with a gestational age interval of 20 to 50 weeks.

have been included in the denominator for the calculation of the percentage of preterm live births. It is evident that the use of DLNM interval data for the calculation of the proportion of post-term live births (42 weeks or longer) would provide only a crude and overestimated index at best. This conclusion is in agreement with those of a recent study comparing the DLNM gestational age interval with ultrasound data (*38*).

These results indicate that nonreporting of gestational age can lead to a marked under-estimation of the percentage of preterm births in a population. Changes in the completeness of reporting of DLNM that are not recognized can lead to misinterpretation of apparent trends in preterm percentages. Imputing data for births with missing DLNM information may appreciably reduce this bias, especially in years with high levels of incomplete reporting. While imputing may increase the error in the estimate of the gestational age interval at the individual case level, this must be weighed against the potential bias created by excluding these cases from assessments of preterm birth in a population. Moreover, these findings emphasize the importance of identifying both the completeness of reporting and the use of any imputation or deletion strategies when employing population-based DLNM data for the calculation of preterm percentages and other gestational age-related indicators, for example, small for gestational age and adequacy of prenatal care utilization (*35*).

These data agree with findings of previous investigations (*29, 30*) and indicate that women with information missing on gestational age are predominant in high-risk sociodemographic subgroups. Cases missing more than the day of LNM were more than twice as likely to be low birth weight compared with the total population (table 4, 'Day 15' method). Cases with complete but implausible DLNM data also demonstrated a higher proportion of low weight births than average. While improvements in reporting completeness and the use of imputation techniques have greatly reduced the proportion of cases with missing gestational age data in this data set, the residual excluded cases, with presumably higher than average levels of preterm delivery, continue to create an under-estimation bias in the calculation of indices of preterm delivery.

The 'Day 15' method of imputing a gestational age for a case with a missing DLNM is computationally less complex than the 'Preceding Case' method and makes no assumptions about the distribution of gestational ages within birth weight categories. Previous reports (*21*), indicating the 'Day 15' method may result in over-estimates of preterm delivery, are supported by these data, particularly from the years with a higher proportion of incomplete reporting (fig. 1). The 'Day 15' method can be a useful tool for research that involves both birth weight and gestational age variables, but its use should be clearly noted and employed with caution.

In addition to requiring substantial computer resources, a shortcoming of the 'Preceding Case' imputation method is that it may fail to identify a plausible gestational age value from a preceding case with the required corresponding month of LNM, birth weight, and race of mother values. This issue becomes more pronounced for births with sociodemographic characteristics associated with missing or implausible DLNM information. Although the criteria for potential preceding cases can be redefined to include only those with plausible gestational age values, this approach requires an accepted definition of plausible. The recommendation for a 20- to 50-week criteria for plausible DLNM gestational interval values is in response to the predicament faced in this study of having no established definition of 'plausible' gestational age data for use in either the calculation of preterm percentages or the imputation of missing data.

Accordingly, it is recommended that the 20- to 50-week criteria be used to identify a preceding case for imputation, when the 'Preceding Case' method is employed. The selection of births with only plausible gestational age values was found to improve slightly the efficiency of this method in terms of recapturing cases with missing data and to decrease slightly the resulting preterm percentages.

As an illustration of the impact of these recommendations, table 5 provides biennial percentages of preterm births by race of mother and imputation groups. Overall, these data indicate little appreciable decrease in the percentage of preterm live births during the 12 years investigated. Although South Carolina infant mortality rates have evinced a steady decline over this period (*39*), much of the decline has been credited to improvements in birth weight-specific mortality (*20, 40*). Racial disparities in preterm delivery have also persisted, so that the risk of a preterm live birth to nonwhite women continues to be more than double that of whites. Since there continues to be less complete reporting among nonwhites, this disparity may be slightly greater than these numbers indicate.

The latest revision of the U.S. standard live birth certificate introduces significant changes from past versions. Its introduction provides a timely opportunity to initiate discussions among involved persons regarding methods to maximize the quality of the data collected. As the use of these computerized data sets for program planning, needs assessment, health status monitoring, evaluation, and resource allocation continues to expand in both public health and health care arenas, efforts spent to maintain and increase the quality of the data represent a fundamental part of perinatal health improvement strategies.

References

1. Andersen, H. F., Johnson, T. R. B., Barclay, M. L., and Flora, J. D.: Gestational age assessment, II. Prediction from combined clinical observations. Am J Obstet Gynecol 140: 770–773 (1981).

2. Lubchenco, L. O.: Assessment of gestational age and development at birth. Pediatr Clin North Am 17: 125–145, (1970).

3. Reid, J.: On the duration of pregnancy in the human female. Lancet i(15): 438–440, April 13, 1850; i(20): 596–599, May 18, 1850; ii(3): 77–81, July 20, 1850.

4. Farr, V., and Mitchell, R. G.: Estimation of gestational age in the newborn infant. Am J Obstet Gynecol 103: 380–383 (1969).

5. Dubowitz, L. M. S., Dubowitz, V, and Goldberg, C.: Clinical assessment of gestational age in the newborn infant. J Pediatr 77: 1–10 (1970).

6. Andersen, H. F., Johnson, T. R. B., Barclay, M. L., and Flora, J. D.: Gestational age assessment, I. Analysis of individual clinical observations. Am J Obstet Gynecol 139: 173–177 (1981).

7. Jimenez, J. M., Tyson, J. E., and Reisch, J. S.: Clinical measures of gestational age in normal pregnancies. Obstet Gynecol 61: 438–443 (1983).

8. Ott, W. J.: Accurate gestational dating. Obstet Gynecol 66: 311–315 (1985).

9. Treloar, A. E., Behn, B. G., and Cowan, D. W.: Analysis of the gestational interval. Am J Obstet Gynecol 99: 34–45 (1967).

10. Frazier, T. M.: Error in reported date of last menstrual period. Am J Obstet Gynecol 77: 915–918 (1959).

11. Gray, H.: Reported date of menstrual period. Western J Surg 70: 31–33, January-February 1962

12. Hotelling, H., and Hotelling, F.: A new analysis of duration of pregnancy data. Am J Obstet Gynecol 23: 643–657 (1932).

13. Karn, M. N.: Length of human gestation with special reference to prematurity. Ann Eugenics 14: 44–59, October 1947.

14. Gibson, J. R., and McKeown, T.: Observations on all births (23,970) in Birmingham, 1947. Br J Soc Med 4: 221–233, October 1950.

15. Gruenwald, P.: Growth of the human fetus, I. Normal growth and its variation. Am J Obstet Gynecol 94: 1112–1119 (1966).

16. Battaglia, F. C., Frazier, T. M., and Hellegers, A. E.: Birth weight, gestational age and pregnancy outcome with special reference to high birth weight-low gestational age infants. Pediatrics 37: 417–422 (1966).

17. Milner, R. D. G., and Richards, B.: An analysis of birth weight by gestational age of infants born in England and Wales, 1967 to 1971. J Obstet Gyn Brit Common 81: 956–967, December 1974.

18. David, R. J.: The quality and completeness of birth weight and gestational age data in computerized birth files. Am J Public Health 79: 964–973 (1980).

19. Williams, R. L., et al. Fetal growth and perinatal viability in California. Obstet Gynecol 59: 624–632 (1982).

20. Alexander, G. R., Tompkins, M. E., Altekruse, J. M., and Hornung, C. A.: Racial differences in the relation of birth weight and gestational age to neonatal mortality. Public Health Rep 100: 539–547, September-October 1985.

21. Taffel, S., Johnson, D., and Heuser, R.: A method of imputing length of gestation on birth certificates. Vital Health Stat [2] No. 93 (1982).

22. Boyce, A., Mayaux, M., and Schwartz, D.: Classical and 'true' gestational postmaturity. Am J Obstet Gynecol 125: 911–914 (1976).

23. Hertz, R. H., et al. Clinical estimation of gestational age: rules for avoiding preterm delivery. Am J Obstet Gynecol 131: 395–402 (1978).

24. Schwartz, S., and West, H.: Potentialities and limitations of medical data of official birth certificates. Am J Public Health 50: 338–345 (1960).

25. Hammes, L. M., and Treloar, A. E.: Gestational interval from vital records. Am J Public Health 60: 1496–1505 (1970).

26. Alexander, G. R., Petersen, D. J., Powell-Griner, E., and Tompkins, M. A.: Comparison of gestational age reporting methods based on physician estimate and date of last normal menses from fetal death reports. Am J Public Health 79: 600–602 (1989).
27. Petersen, D. J., Alexander, G. R., Powell-Griner, E., and Tompkins M.: Variations in the reporting of gestational age at induced termination of pregnancy. Am J Public Health 79: 603–606 (1989).
28. Wenner, W. H., and Young, E. B.: Nonspecific date of last menstrual period: an indication of poor reproductive outcome. Am J Obstet Gynecol 120: 1071–1079 (1974).
29. Buekens, P., Delvoye, P., Wollast, E., and Robyn, C.: Epidemiology of pregnancies with unknown last menstrual period. J Epidemiol Community Health 38: 79–80 (1984).
30. Hall, M. N., et al: The extent and antecedents of uncertain gestation. Brit J Obstet Gynaecol 92: 445–451 (1985).
31. Hoffman, H. J., Stark, C. R., Lundin, F. E., and Ashbrook, J. D.: Analysis of birth weight, gestational age, and fetal viability, U.S. births, 1968. Obstet Gynecol Surv 29: 651–681 (1974).
32. Hall, M. H., and Carr-Hill, R. A.: The significance of uncertain gestation for obstetric outcome. Brit J Obstet Gynaecol 92: 452–460 (1985).
33. David, R. J.: Population-based intrauterine growth curves from computerized birth certificates. South Med J 76: 1401–1406 (1983).
34. Binkin, N. J., Williams, R. L., Hogue, C. J. R., and Chen, P. M.: Reducing black neonatal mortality. Will improvement in birth weight be enough? JAMA 253: 372–375, Jan. 18, 1985.
35. Alexander, G. R., and Cornely, D. A.: Prenatal care utilization: Its measurement and relationship to pregnancy outcome. Am J Prev Med 35: 242–253, September-October 1987.
36. Silverman, W. A., et al: Nomenclature for duration of gestation, birth weight and intrauterine growth. Pediatrics 39: 935–939 (1967).
37. Kleinman, J. C., and Kessel, S. S.: Racial differences in low birth weight. Trends and risk factors. New Engl J Med 317: 749–753 (1987).
38. Kramer, M. S., McLean, F. H., Boyd, M. E., and Usher, R. H.: The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. JAMA 260: 3306–3308 (1988).
39. The Division of Biostatistics, Office of Vital Records and Public Health Statistics, South Carolina Department of Health and Environmental Control: South Carolina vital and morbidity statistics 1985, vol I: Annual Vital Statistics Series 1987.
40. Tompkins, M. E., et al.: The risk of low birth weight: alternative models of neonatal mortality. Am J Epidemiol 122: 1067–1079 (1985).