

DATA HEALTH STATISTICS

A Method of Imputing Length of Gestation on Birth Certificates

A method of imputing the period of gestation is outlined for births when the month and year of the date of the last menstrual period are reported on the birth certificate but the day is missing. This method uses the recorded birth weight and months of gestation.

Data Evaluation and Methods Research Series 2, No. 93

DHHS Publication No. (PHS) 82-1367

U.S. Department of Health and Human Services
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Md.
May 1982

SUGGESTED CITATION

National Center for Health Statistics, S. Taffel, D. Johnson, and R. Heuser: A method of imputing length of gestation on birth certificates. *Vital and Health Statistics. Series 2, No. 93.* DHHS. Pub. No. (PHS) 82-1367. Public Health Service. Washington. U.S. Government Printing Office, May 1982.

Library of Congress Cataloging in Publication Data

Taffel, Selma.

A method of imputing length of gestation on birth certificates.

(Data evaluation and methods research, series 2 ; no. 93) (DHHS publications ; no. (PHS) 82-1367)

Prepared by Selma Taffel, David Johnson, and Robert Heuser.

"May 1982"—T.p. verso.

Includes bibliographical references.

1. Conception—Date of. 2. Pregnancy—Duration—Tables. 3. Pregnancy—United States—Duration—Statistics. I. Johnson, David P. II. Heuser, Robert L. III. National Center for Health Statistics (U.S.). IV. Title. V. Series: Vital and health statistics. Series 2, Data evaluation and methods research ; no. 93. VI. Series: DHHS publication ; no. (PHS) 82-1367.

RA409.U45 no. 93 [RA1138] 312'.0723s 82-600076
ISBN 0-8406-0251-0 AACR2

National Center for Health Statistics

DOROTHY P. RICE, *Director*

ROBERT A. ISRAEL, *Deputy Director*

JACOB J. FELDMAN, Ph.D., *Associate Director for Analysis and Epidemiology*

GAIL F. FISHER, Ph.D., *Associate Director for the Cooperative Health Statistics System*

GARRIE J. LOSEE, *Associate Director for Data Processing and Services*

ALVAN O. ZARATE, Ph.D., *Assistant Director for International Statistics*

E. EARL BRYANT, *Associate Director for Interview and Examination Statistics*

ROBERT C. HUBER, *Associate Director for Management*

MONROE G. SIRKEN, Ph.D., *Associate Director for Research and Methodology*

PETER L. HURLEY, *Associate Director for Vital and Health Care Statistics*

ALICE HAYWOOD, *Information Officer*

Vital and Health Care Statistics Program

PETER L. HURLEY, *Associate Director*

GLORIA KAPANTAIS, *Assistant to the Director for Data Policy, Planning, and Analysis*

Division of Vital Statistics

JOHN E. PATTERSON, *Director*

ALICE M. HETZEL, *Deputy Director*

ROBERT L. HEUSER, M.A., *Chief, Natality Statistics Branch*

MABEL G. SMITH, *Chief, Statistical Resources Branch*

JOSEPH D. FARRELL, *Chief, Computer Applications Staff*

Contents

Introduction	1
Methodology	2
Imputation procedures	4
Discussion	7
 Appendix	
Contents	9
Technical notes	10
 List of text figures	
1. Percent distribution of live births by period of gestation for records with complete date of last menstrual period and for these records plus missing-day records with day imputed as the 15th: Total of 47 reporting States and the District of Columbia, 1978	4
2. Percent distribution of live births by period of gestation for records with complete date of last menstrual period with originally reported day and with the 15th as day: Total of 47 reporting States and the District of Columbia, 1978	4
3. Percent distribution of live births by period of gestation for records with complete date of last menstrual period with originally reported day and with a randomly assigned day: Total of 47 reporting States and the District of Columbia, 1978	6
 List of text tables	
A. Percent of live-birth certificates with incomplete reporting of date of last menstrual period, by race: Total of reporting areas, 1968-78	1
B. Percent distribution of live births in 20-percent sample file and in complete file by period of gestation: Total of 47 reporting States and the District of Columbia, 1978	2
C. Number and percent distribution of live births by selected characteristics, according to reporting of date of last menstrual period: Total of reporting areas, 1978	3
D. Percent distribution of live births with original gestation of 39, 40, and 41 weeks by weeks of gestation recomputed with the 15th as day of last menstrual period: Total of 47 reporting States and the District of Columbia, 1978	5
E. Percent of live births with selected weeks of gestation based on alternate methods of computing period of gestation	6
F. Percent distribution of live births by period of gestation for records with complete date of last menstrual period, imputed missing-day records, and merged file: Total of 47 reporting States and the District of Columbia, 1978	7

Symbols

- - - Data not available
 - . . . Category not applicable
 - Quantity zero
 - 0.0 Quantity more than zero but less than 0.05
 - Z Quantity more than zero but less than 500 where numbers are rounded to thousands
 - * Figure does not meet standards of reliability or precision
-

A Method of Imputing Length of Gestation on Birth Certificates

by Selma Taffel, David Johnson, and Robert Heuser, M.A.,
Division of Vital Statistics

Introduction

The length of gestation of a newborn infant is defined as beginning with the first day of the last normal menstrual period and ending with the day of birth. Until 1968, the birth certificates of almost all States asked for period of gestation in terms of weeks or months. This method of reporting resulted in a substantial heaping of births at 40 weeks of gestation because the gestational period of apparently full term births was frequently reported as 9 months or 40 weeks. To minimize errors in reporting, the 1968 revision of the U.S. Standard Certificate of Live Birth asked for the date the last normal menses began as the basis for the computation of length of gestation. As is shown in the appendix, in 1968 this question was included on the live-birth certificates of 36 States and the District of Columbia; by 1978 the reporting area had expanded to 47 States and the District of Columbia.

At present, the National Center for Health Sta-

tistics derives the length of gestation from the reported month, day, and year of the last menstrual period (LMP). Since 1968, however, in areas reporting that information, the proportion of birth certificates having one or more of these three elements missing has ranged from 18 to 20 percent for white births and from 24 to 28 percent for black births. Between 12 and 16 percent of all records were missing only the day of LMP (table A). Thus, a substantial portion of the incomplete reporting (75 percent in 1978) is due to the absence of only the day of LMP. The National Center for Health Statistics in the past has made no attempt to derive gestational data for records with incomplete reporting. This report examines the feasibility of imputing the length of gestation for records on which the day of LMP is missing, but the month and year are known. Three methods of imputation are examined. The method chosen for implementation introduces the least bias, since it uses both reported months of gestation and birth weight.

Table A. Percent of live-birth certificates with incomplete reporting of date of last menstrual period, by race: Total of reporting areas, 1968-78

Year	Day, month, or year missing			Day only missing		
	All races ¹	White	Black	All races ¹	White	Black
1978	19.9	18.4	26.8	14.9	13.5	20.5
1977	20.6	19.1	27.8	15.9	---	---
1976	21.1	19.6	28.0	16.4	---	---
1975	21.0	19.6	27.2	15.9	---	---
1974	20.8	19.5	27.5	15.7	---	---
1973	20.7	19.3	27.4	15.5	---	---
1972	20.0	18.6	26.8	15.0	---	---
1971	19.6	18.3	26.1	14.2	---	---
1970	19.6	18.4	25.5	13.7	---	---
1969	19.1	18.1	24.4	13.6	---	---
1968	20.3	19.3	² 25.2	12.0	---	---

¹Includes races other than white and black.

²Represents all races other than white; information not available for black births separately.

Methodology

A 20-percent systematic sample was drawn of all 1978 live-birth certificates for mothers residing in areas reporting the date of the last normal menstrual period (LMP). As shown in table B, the distribution of births by period of gestation of the 20-percent sample is a very close approximation of the complete file from which it was drawn. As a first step, records where month and year of LMP were reported, but day was missing (designated hereafter as "missing-day" records or births), were evaluated to see if they represented an atypical group of births with respect to demographic and health-related measures that are highly correlated with gestational period. If the missing-day records differed substantially from records with complete date, then the omission of such records from computation of gestational period would introduce a bias of unknown proportion.

In fact, major differences were found for a number of characteristics examined (table C). The proportion of births that were black was substantially higher for missing-day records than for records where the

date was completely reported (23.4 percent compared with 15.3 percent). There was a far higher proportion of teenage mothers for whom the day of LMP was missing (22.1 percent) than for whom there was complete reporting (15.2 percent). Consistent with this finding was the generally lower educational attainment of mothers with incomplete reporting. About one-third (33.7 percent) had not completed high school compared with about one-quarter (24.3 percent) of the mothers where the LMP date was complete. There was a greater likelihood that the mother was unmarried when the day was missing (24.0 percent) than when the date was complete (14.6 percent). The incidence of low birth weight (2,500 grams or less) was more than one-third higher among the missing-day births (9.0 percent compared with 6.6 percent); and mothers were more likely to delay prenatal care to the last trimester of pregnancy or to have no care in the missing-day group (7.3 percent compared with 4.5 percent). Thus there is a profile of a group of relatively "high risk" births, for

Table B. Percent distribution of live births in 20-percent sample file and in complete file by period of gestation: Total of 47 reporting States and the District of Columbia, 1978

Weeks of gestation	All races ¹		White		Black	
	20-percent sample file	Complete file	20-percent sample file	Complete file	20-percent sample file	Complete file
Total	100.0	100.0	100.0	100.0	100.0	100.0
Under 20 weeks	0.0	0.0	0.0	0.0	0.1	0.1
20-27 weeks	0.5	0.5	0.4	0.4	1.0	1.0
28-31 weeks	0.8	0.8	0.6	0.6	1.6	1.6
32-35 weeks	3.5	3.5	3.0	2.9	5.9	5.9
36 weeks	2.4	2.3	2.1	2.1	3.3	3.3
37-39 weeks	28.7	28.7	28.3	28.3	29.3	29.5
40 weeks	17.8	17.9	18.7	18.8	13.6	13.5
41-42 weeks	19.7	19.7	21.3	21.3	12.9	12.7
43 weeks and over	6.8	6.8	7.1	7.1	5.7	5.5
Not stated	19.8	19.9	18.4	18.4	26.7	26.8

¹Includes races other than white and black.

Table C. Number and percent distribution of live births by selected characteristics, according to reporting of date of last menstrual period: Total of reporting areas, 1978

[Based on a 20-percent sample of births]

Selected characteristics	All races ¹		White		Black	
	Complete reporting of date	Day missing	Complete reporting of date	Day missing	Complete reporting of date	Day missing
Age of mother						
Number	2,434,370	441,315	1,984,795	325,015	373,600	103,135
Percent	100.0	100.0	100.0	100.0	100.0	100.0
Under 15 years	0.2	0.6	0.1	0.3	0.8	1.7
15-19 years	14.9	21.5	13.1	18.0	25.6	32.7
20-24 years	34.0	34.2	34.0	34.3	35.8	34.3
25-29 years	31.5	26.7	33.0	29.0	23.1	19.0
30-34 years	14.7	12.7	15.3	14.0	10.3	8.4
35-39 years	3.9	3.6	3.9	3.7	3.5	3.1
40-49 years	0.7	0.7	0.7	0.7	0.9	0.7
Years of school completed by mother						
Number ²	2,387,715	431,225	1,942,450	316,180	371,990	102,625
Percent	100.0	100.0	100.0	100.0	100.0	100.0
0-8 years	5.3	7.0	5.3	7.0	5.0	6.6
9-11 years	19.0	26.7	16.7	22.7	31.2	39.2
12 years	43.8	41.7	44.5	42.9	42.1	39.0
13-15 years	18.1	14.7	18.6	15.7	15.4	11.4
16 years or more	13.8	9.9	14.9	11.7	6.3	3.8
Marital status of mother						
Number	1,566,080	301,920	1,283,760	223,555	237,955	69,360
Percent	100.0	100.0	100.0	100.0	100.0	100.0
Married	85.4	76.0	92.3	87.3	48.4	39.7
Unmarried	14.6	24.0	7.7	12.7	51.6	60.3
Birth weight						
Number ²	2,434,370	441,315	1,984,795	325,015	373,600	103,135
Percent	100.0	100.0	100.0	100.0	100.0	100.0
500 grams or less	0.1	0.1	0.1	0.1	0.2	0.2
501-1,000 grams	0.4	0.6	0.3	0.4	0.8	1.0
1,001-1,500 grams	0.6	0.9	0.5	0.7	1.2	1.4
1,501-2,000 grams	1.3	1.8	1.1	1.4	2.3	3.1
2,001-2,500 grams	4.3	5.7	3.7	4.7	7.3	9.0
2,501-3,000 grams	16.4	18.8	14.8	16.4	24.6	26.1
3,001-3,500 grams	37.4	37.0	37.0	36.9	38.6	37.1
3,501-4,000 grams	29.0	26.1	30.9	28.8	19.9	17.6
4,001-4,500 grams	8.8	7.5	9.7	8.7	4.3	3.8
4,501-5,000 grams	1.6	1.3	1.8	1.5	0.7	0.6
5,001 grams or more	0.2	0.2	0.2	0.3	0.1	0.1
2,500 grams or less	6.6	9.0	5.6	7.3	11.8	14.7
Month of pregnancy prenatal care began						
Number ²	2,434,370	441,315	1,984,795	325,015	373,600	103,135
Percent	100.0	100.0	100.0	100.0	100.0	100.0
1st or 2d month	49.7	47.8	52.5	51.9	36.4	35.4
3d month	26.6	23.1	26.9	23.4	25.0	22.2
4th-6th month	19.2	21.8	17.0	18.8	30.3	31.2
7th-9th month	3.7	4.8	3.1	4.0	6.4	6.9
No prenatal care	0.8	2.5	0.6	1.9	2.0	4.3

¹Includes races other than white and black.

²Includes records with information not stated, which have been distributed.

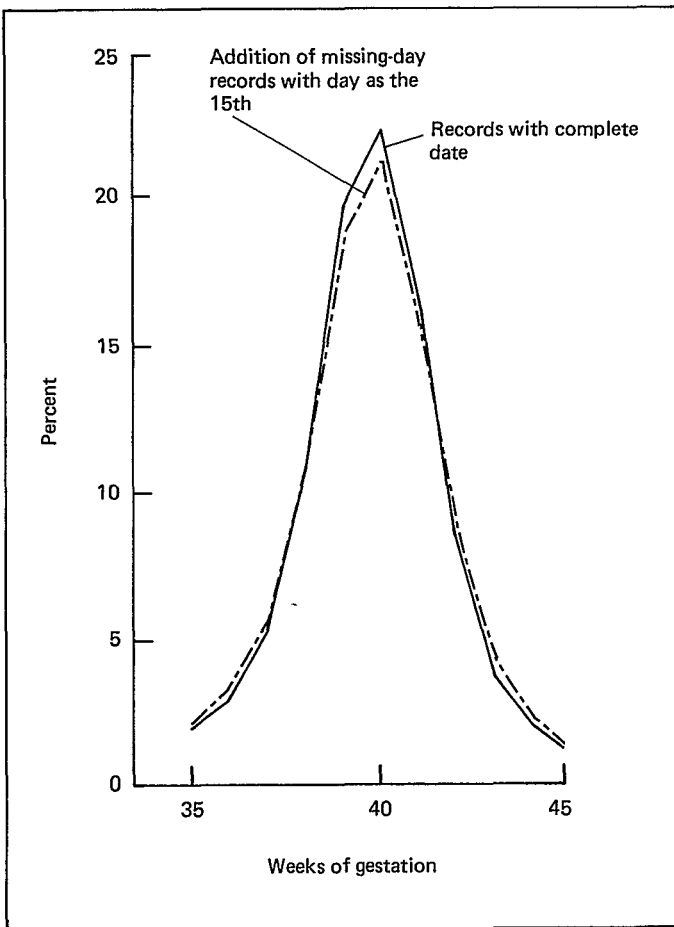


Figure 1. Percent distribution of live births by period of gestation for records with complete date of last menstrual period and for these records plus missing-day records with day imputed as the 15th: Total of 47 reporting States and the District of Columbia, 1978

[Based on a 20-percent sample of births]

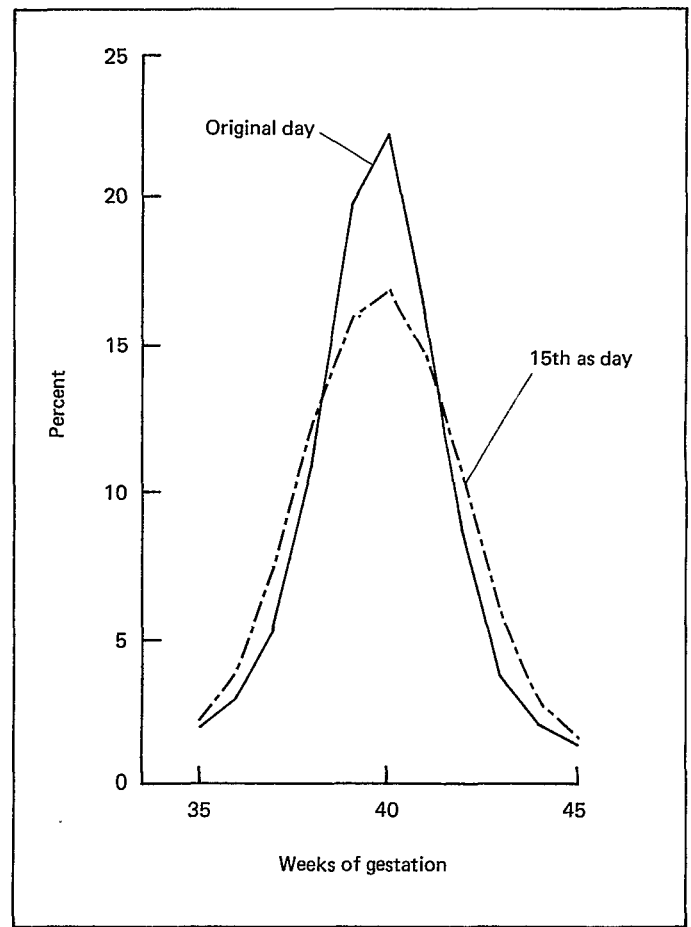


Figure 2. Percent distribution of live births by period of gestation for records with complete date of last menstrual period with originally reported day and with the 15th as day: Total of 47 reporting States and the District of Columbia, 1978

[Based on a 20-percent sample of births]

which there is an increased likelihood of short gestational periods. Since these births are shown in the "not stated" gestation category, the published data with known gestation are unquestionably biased towards longer periods of gestation.

Imputation procedures

The ideal imputation procedure would eliminate the bias caused by the exclusion of missing-day records, while not introducing other biases due to the method of imputation. Three methods were considered: Two involved an arbitrary designation of the day of LMP, while the third was based on the relationship between months of gestation and birth weight.

Assignment of the 15th as day of LMP.—The first method, assignment of the 15th as day of LMP for all missing-day records, resulted in an expanded file of records with relatively fewer births at 39, 40, and 41 weeks of gestation and the same or higher proportion of births at longer and shorter gestational intervals (figure 1). An indication that the imputation pro-

cedure itself is partly responsible for these differences is evident from a comparison of the distributions shown in figure 2. Both distributions use the records that have complete reporting of date of LMP. The more peaked curve reflects gestational periods derived from the originally reported day, month, and year, while the flatter curve is the one that would result if the original day of LMP were replaced in all records by the 15th day of the month. The decrease in proportion of records at 40 weeks of gestation from 22.2 percent to 16.9 percent is thus due solely to the imputation procedure.

The reason can be seen by examining the procedure more closely. When the day of LMP is changed to 15, there is a maximum change of 3 weeks in the length of gestation. For example, for records with an original gestation of 40 weeks, the recomputed gestation ranged from 38 to 43 weeks, as shown in table D. The recomputed distributions were similar for the original gestational ages of 39, 40, and 41 weeks. The length of gestation did not change for about 25 percent of the records; there was a change of 1 week for about 45 percent (half increased and half decreased),

Table D. Percent distribution of live births with original gestation of 39, 40, and 41 weeks by weeks of gestation recomputed with the 15th as day of last menstrual period: Total of 47 reporting States and the District of Columbia, 1978

[Based on a 20-percent sample of births]

Recomputed week of gestation	Original week of gestation					
	39 weeks	40 weeks	41 weeks	39 weeks	40 weeks	41 weeks
	Percent			Number		
Total	100.0	100.0	100.0	477,495	541,315	392,705
37 weeks	13.3	-	-	63,505	-	-
38 weeks	22.0	14.8	-	105,175	80,130	-
39 weeks	25.6	22.4	16.0	122,215	121,355	62,975
40 weeks	22.4	25.0	22.4	106,975	135,325	87,825
41 weeks	15.8	22.4	24.8	75,670	121,455	97,365
42 weeks	0.8	14.7	22.6	3,955	79,600	88,890
43 weeks	-	0.6	13.7	-	3,450	53,735
44 weeks	-	-	0.5	-	-	1,915

and a change of 2-3 weeks for 30 percent of the records (half increased and half decreased).

From these percents it appears that there is an even exchange between adjacent weeks when gestation is recomputed. For example, 22.4 percent of the original 40-week records changed to 39 weeks, and the same percent of the original 39-week records changed to 40 weeks. Although the percents were the same, the numbers of records changed were not the same because the original distribution has a peak at 40 weeks, with fewer records at 39 and 41 weeks. There were more changed to 39 weeks (121,355) than were changed to 40 weeks (106,975). A similar exchange situation exists between 40 and 41 weeks, with more records changed to 41 than to 40 weeks.

Differences in the distributions shown in figure 1 are thus due both to the bias introduced by the imputation procedure and to the shorter gestational periods of the missing-day births. An indication of the contribution of the latter component was derived by using birth weight information recorded on the missing-day and completely reported records. For each 500-gram weight interval, a percent distribution of completely reported records by weeks of gestation was calculated. These were used as models to distribute missing-day records having comparable birth weights by weeks of gestation. A composite weighted distribution by weeks of gestation was then derived by combining the distributions of the reported and missing-day records (table E, column C). Differences between this combined distribution and the original distribution based only on fully reported records (column A) are solely a reflection of the generally lower birth weights of the missing-day records.

Table E presents a comparison of these sets for the most frequently reported weeks of gestation and for all gestational periods of under 37 weeks combined. Referring again to figure 1, and using the information in table E, it can be demonstrated that most of the difference between the original distribution (column A) and the distribution that incorpo-

rates missing-day records with the 15th imputed as day of LMP (column B) is due to the imputation process itself (column B minus column C). For example, 91 percent of the difference in the proportion of records assigned to the modal group of 40 weeks is attributable to the imputation procedure and only 9 percent to differences in birth weight. As would be expected, among premature births (under 37 weeks of gestation) the generally lower birth weight of the missing-day records gains importance as a contributing factor. In this example, 33 percent of the difference in the proportion of records with under 37 weeks gestation is due to the birth weight factor and 67 percent to the imputation procedure. It appears, then, that imputing the 15th day of the month as day of LMP introduces substantial bias due to the imputation procedure itself.

Random assignment of day of LMP.—The second method of imputation investigated was the random assignment of a day of LMP. To assess the validity of this procedure, the original day of LMP of all completely reported records was replaced by a randomly numbered day from 1 to 28, 1 to 30, or 1 to 31, as appropriate for the reported month. The resulting distribution is considerably flatter than when the imputed day is the 15th of the month (figures 2 and 3), an indication that even greater bias is introduced by this method than by the first method. This is because the maximum change in the length of gestation when the day is assigned randomly is 5 weeks, compared with 3 weeks for the first method.

Imputation based on months of gestation and birth weight.—Since the arbitrary designation of a day of LMP resulted in unacceptable distortions of the original data, a third method was tested that bypassed the day of LMP and relied instead on the relationship between months of gestation and birth weight. The reporting of birth weight on birth certificates is nearly universal. In 1978 only 0.3 percent of the birth certificates in the LMP reporting area lacked valid information on newborn weight. Weeks of gesta-

Table E. Percent of live births with selected weeks of gestation based on alternate methods of computing period of gestation

[Based on a 20-percent sample of births]

Selected weeks of gestation	Records with complete date of last menstrual period	Records with complete date of last menstrual period plus missing-day records	
	Gestation based on original reporting of date	Gestation imputed for missing-day records by assigning the 15th as day	Gestation for missing-day records distributed according to birth weight
	(A)	(B)	(C)
35 weeks	1.9	2.0	1.9
36 weeks	2.9	3.2	3.0
37 weeks	5.2	5.6	5.3
38 weeks	10.9	11.0	10.9
39 weeks	19.6	18.8	19.6
40 weeks	22.2	21.1	22.1
41 weeks	16.1	15.7	16.0
42 weeks	8.5	8.6	8.4
43 weeks	3.9	4.2	3.9
44 weeks	2.0	2.2	2.0
45 weeks	1.1	1.2	1.1
Under 37 weeks	8.9	9.8	9.2

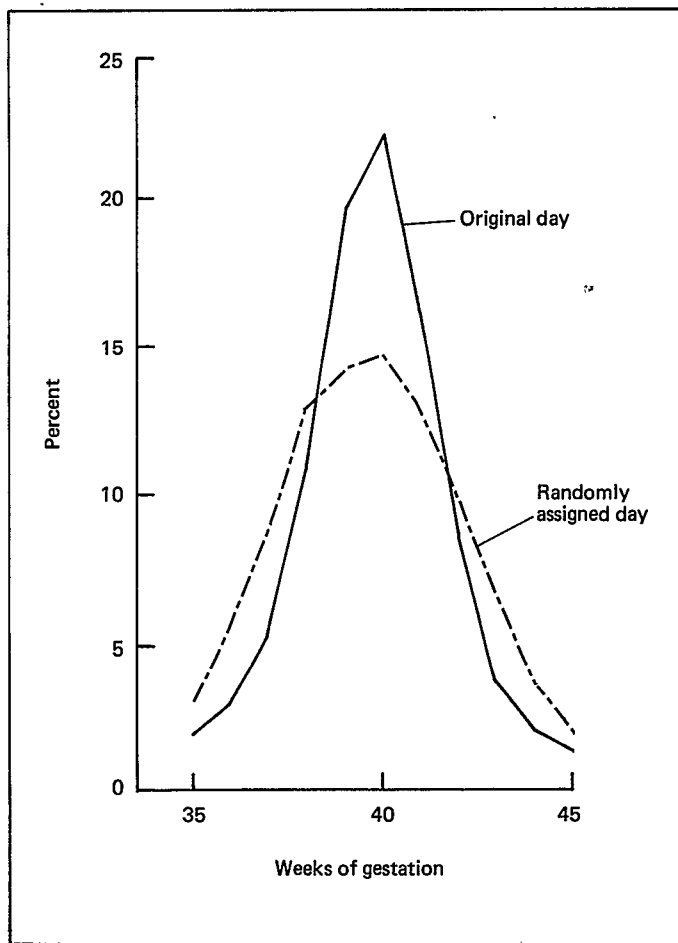


Figure 3. Percent distribution of live births by period of gestation for records with complete date of last menstrual period with originally reported day and with a randomly assigned day: Total of 47 reporting States and the District of Columbia, 1978

[Based on a 20-percent sample of births]

tion for missing-day records can be assigned using the weeks of gestation of completely reported records with the same computed months of gestation and same birth weight. (The number of months of gestation was computed for both missing-day records and completely reported records by subtracting the reported month of LMP from the reported month of birth.)

To avoid any bias in the assignment procedure, the total file of records was sorted into consecutive certificate number order within each State. Each missing-day record was then assigned the gestational period in weeks of the preceding completely reported record having the same computed months of gestation and the same 500-gram birth weight interval. Records with 10 or more months of reported gestation were all considered as having a gestational age of 10 months in this procedure since it is likely that gestational periods of 11 or 12 months are erroneous. Incorrect reporting of the month of LMP can occur as a result of menstrual irregularities. For women using oral contraceptives, the time between the last menstrual period and ovulation may be prolonged after withdrawal from usage, and the derived period of gestation artificially lengthened (L. O. Lubchenco: *The High Risk Infant*. Philadelphia. W. B. Saunders Company, 1976. p. 10).

Table F shows the distribution by weeks of gestation of completely reported records, missing-day records with gestation imputed using the procedure just outlined, and the expanded distribution that incorporates the imputed missing-day records. The effect of including these imputed missing-day records is to increase the proportion of premature births from 8.9 to 9.4 percent and to lower slightly the proportion of

Table F. Percent distribution of live births by period of gestation for records with complete date of last menstrual period, imputed missing-day records, and merged file: Total of 47 reporting States and the District of Columbia, 1978

[Based on a 20-percent sample of births]

<i>Weeks of gestation</i>	<i>Records with complete date</i>	<i>Missing-day records with gestation imputed using month of last menstrual period and birth weight</i>	<i>Merged file</i>
Total	100.0	100.0	100.0
17 weeks	0.0	0.0	0.0
18 weeks	0.0	0.1	0.0
19 weeks	0.0	0.0	0.0
20 weeks	0.0	0.1	0.0
21 weeks	0.0	0.1	0.0
22 weeks	0.1	0.1	0.1
23 weeks	0.1	0.1	0.1
24 weeks	0.1	0.2	0.1
25 weeks	0.1	0.2	0.1
26 weeks	0.1	0.2	0.1
27 weeks	0.1	0.2	0.2
28 weeks	0.2	0.2	0.2
29 weeks	0.2	0.4	0.2
30 weeks	0.3	0.5	0.3
31 weeks	0.4	0.5	0.4
32 weeks	0.5	0.8	0.6
33 weeks	0.8	1.1	0.8
34 weeks	1.2	1.6	1.2
35 weeks	1.9	2.4	2.0
36 weeks	2.9	3.5	3.0
37 weeks	5.2	5.8	5.3
38 weeks	10.9	10.8	10.9
39 weeks	19.6	18.3	19.4
40 weeks	22.2	20.8	22.0
41 weeks	16.1	15.3	16.0
42 weeks	8.5	8.1	8.4
43 weeks	3.9	4.0	3.9
44 weeks	2.0	2.0	2.0
45 weeks	1.1	1.1	1.1
46 weeks	0.6	0.6	0.6
47 weeks	0.3	0.4	0.4
48 weeks	0.2	0.2	0.2
49 weeks	0.2	0.2	0.2
50 weeks	0.1	0.1	0.1
51 weeks	0.1	0.1	0.1
52 weeks	0.1	0.1	0.1
Under 37 weeks	8.9	12.1	9.4

births at 39, 40, 41, and 42 weeks of gestation. These changes seem reasonable in light of the atypical demographic profile and generally lower birth weight of missing-day births.

Discussion

The length of gestation in weeks cannot be ascertained from information reported on a substantial number of live-birth certificates each year because the day of LMP is not reported. In 1978, 15 percent of all birth certificates in the areas that included a question on date of LMP had month and year of LMP entered, but lacked information on the day of LMP. The demographic and health characteristics of these births indicate that they are a relatively "high-risk" group, with lower than average birth weight. Accord-

ingly it is expected that their gestational periods would be shorter than average. It is estimated that classifying the gestational period of these births as "not stated" has the effect of understating the true proportion of premature births by about 5 percent.

Exclusion of these records also affects comparisons between period of gestation and other demographic variables. For example, before the addition of the imputed missing-day records, 27.0 percent of very short gestation births (less than 28 weeks) were to teenage mothers; after inclusion of these records, the proportion of teenage mothers increased by 7 percent, to 29.0 percent of all mothers. Similarly, the proportion of very short gestation births to mothers who had not completed high school rose by 7 percent (from 35.4 percent to 37.8 percent).

A commonly used method of deriving weeks of

gestation when the day of LMP is missing is to assume the 15th to be the day of LMP. This study demonstrates that this procedure introduces noticeable bias. The random assignment of a day of LMP from 1 to 31 results in even greater distortions in period of gestation data. An alternative method of imputation is outlined that makes use of the recorded birth weight and month of gestation and that introduces a maximum imputed error of 3 weeks in the length of gestation. This procedure corrects for the previous understatement of premature births by reducing the level of unknown gestation while apparently intro-

ducing only low levels of procedural bias. It also preserves any incompatibility between birth weight and period of gestation to the extent this is inherent in the data.

The National Center for Health Statistics plans to include this imputation procedure in future revisions of its processing operations. Imputation will be performed independently for white births, black births, and births of other races, since the relationship between birth weight and gestation is substantially different among these racial groups.

Appendix

Contents

Technical notes	10
Appendix table	
States reporting date last normal menstrual period began: United States, 1968-78	11

Appendix. Technical notes

The period of gestation is defined as beginning with the first day of the last normal menstrual period (LMP) and ending with the day of birth. The LMP date is used as the initial date since it can be more accurately determined than the date of conception, which usually occurs about 2 weeks after the LMP date. Births occurring prior to 37 weeks of gestation are considered to be preterm or premature for purposes of classification. This distinction is in accordance with the one adopted by the World Health Organization Expert Group on Prematurity established in 1950.

The data presented in this report are derived from a 20-percent sample of 1978 birth certificates of

States that included a question on date of last normal menstrual period. The appendix table shows the areas requesting the date of LMP for the years 1968-78. The sample design was systematic, with birth certificates having certificate numbers ending in a 2 or 6 selected for inclusion.

As discussed in the text and shown in table B, the birth certificates chosen for this study are representative of the entire file of births in terms of the distribution by period of gestation. All aggregate numbers shown in this report are based on observed totals of the sample records; percents shown are presented as estimates for the underlying population.

Appendix table. States reporting date last normal menstrual period began: United States, 1968-78

State	1978	1977	1976	1975	1974	1973	1972	1971	1970	1969	1968
Alabama	X	X	X								
Alaska	X	X	X	X	X	X	X	X	X	X	X
Arizona	X	X	X	X	X	X	X	X	X	X	X
Arkansas	X										
California	X	X	X	X	X	X	X	X	X	X	X
Colorado	X	X	X	X	X	X	X	X	X	X	X
Connecticut											
Delaware	X	X	X	X	X	X	X				
District of Columbia	X	X	X	X	X	X	X	X	X	X	X
Florida	X	X	X	X	X	X	X	X	X		
Georgia	X	X	X	X	X	X					
Hawaii	X	X	X	X	X	X	X	X	X	X	X
Idaho	X										
Illinois	X	X	X	X	X	X	X	X	X	X	X
Indiana	X	X	X	X	X	X	X	X	X	X	X
Iowa	X	X	X	X	X	X	X	X	X	X	X
Kansas	X	X	X	X	X	X	X	X	X	X	X
Kentucky	X	X	X	X	X	X	X	X	X	X	X
Louisiana	X	X	X	X	X	X	X	X	X	X	X
Maine	X	X	X	X	X	X	X	X	X	X	X
Maryland	X	X	X	X	X	X	X	X	X	X	X
Massachusetts	X	X	X							X	
Michigan	X	X	X	X	X	X	X	X	X	X	X
Minnesota	X	X	X	X	X	X	X	X	X	X	X
Mississippi	X	X	X	X	X	X	X	X	X	X	X
Missouri	X	X	X	X	X	X	X	X	X	X	X
Montana	X	X	X	X	X	X	X	X	X	X	X
Nebraska	X	X	X	X	X	X	X	X	X	X	X
Nevada	X	X	X	X	X	X	X	X	X	X	X
New Hampshire	X	X	X	X	X	X	X	X	X	X	X
New Jersey	X	X	X	X	X	X	X	X	X	X	X
New Mexico											
New York	X	X	X	X	X	X	X	X	X	X	X
North Carolina	X	X	X	X	X	X	X	X	X	X	X
North Dakota	X	X	X	X	X	X	X	X	X	X	X
Ohio	X	X	X	X	X	X	X	X	X	X	X
Oklahoma	X	X	X	X	X	X	X	X	X	X	X
Oregon	X	X	X	X	X	X	X	X	X		
Pennsylvania	X										
Rhode Island	X	X	X	X	X	X	X	X	X	X	X
South Carolina	X	X	X	X	X	X	X	X	X	X	X
South Dakota	X	X	X	X	X	X	X	X	X	X	X
Tennessee	X	X	X	X	X	X	X	X	X	X	X
Texas											
Utah	X	X	X	X	X	X	X	X	X	X	X
Vermont	X	X	X	X	X	X	X	X	X	X	X
Virginia	X										
Washington	X	X	X	X	X	X	X	X	X	X	X
West Virginia	X	X	X	X	X	X	X	X	X	X	X
Wisconsin	X										
Wyoming	X	X	X	X	X	X	X	X	X	X	X

Vital and Health Statistics series descriptions

- SERIES 1. Programs and Collection Procedures.**—Reports describing the general programs of the National Center for Health Statistics and its offices and divisions and the data collection methods used. They also include definitions and other material necessary for understanding the data.
- SERIES 2. Data Evaluation and Methods Research.**—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory.
- SERIES 3. Analytical and Epidemiological Studies.**—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.
- SERIES 4. Documents and Committee Reports.**—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.
- SERIES 10. Data from the National Health Interview Survey.**—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in the continuing national household interview survey.
- SERIES 11. Data From the National Health Examination Survey and the National Health and Nutrition Examination Survey.**—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.
- SERIES 12. Data From the Institutionalized Population Surveys.**—Discontinued in 1975. Reports from these surveys are included in Series 13.
- SERIES 13. Data on Health Resources Utilization.**—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.
- SERIES 14. Data on Health Resources: Manpower and Facilities.**—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.
- SERIES 15. Data From Special Surveys.**—Statistics on health and health-related topics collected in special surveys that are not a part of the continuing data systems of the National Center for Health Statistics.
- SERIES 20. Data on Mortality.**—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.
- SERIES 21. Data on Natality, Marriage, and Divorce.**—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.
- SERIES 22. Data From the National Mortality and Natality Surveys.**—Discontinued in 1975. Reports from these sample surveys based on vital records are included in Series 20 and 21, respectively.
- SERIES 23. Data From the National Survey of Family Growth.**—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a periodic survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to:

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service
Hyattsville, Md. 20782

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HHS
HHS 396

THIRD CLASS



HRST

From the Office of Health Research, Statistics, and Technology
DHHS Publication No. (PHS) 82-1367, Series 2, No. 93

For listings of publications in the VITAL AND HEALTH STATISTICS series, call 301-436-NCHS