BIRTH DATA QUALITY TECHNICAL NOTES

Washington State Department of Health

Center for Health Statistics, July 2016

Note: Not all states have implemented the 2003 revisions to the birth certificate. For this reason, data for most of the new items (such as multiple race data or maternal morbidity) will be missing for those residents who have their baby in a state which hasn’t yet made the change.

The birth certificate data collection process is not a well-defined entity. Most hospitals use a series of worksheets to collect data from the mother, the medical records, and the physician. However, the type of worksheet used and the method of collecting data from these sources varies from hospital to hospital, as does the amount of follow up used to collect missing information. Starting in 2003, more hospitals use the standardized worksheet provided by the Department of Health.

The method for entering birth certificate data into the computer has changed:

1. 1980-1991: Hospitals or birth attendants complete a paper certificate form and send it to the local health jurisdiction, which then sends it to the state Department of Health, where the data are coded (if necessary) and keyed.
2. 1992: Hospitals or birth attendants begin to use an electronic birth certificate program called the Delivery Certificate Tracker (DCT) to enter birth certificate records, which are then sent directly to the Department of Health. Computer programs now electronically code many items.
3. Mid-1996: First revision of the DCT
4. 1999: Second revision of the DCT
5. 2003: A new web-based system for entering the data replaces the DCT. This system is called the Birth Record Real Time Registration System (BR3). The system has more data edits and makes it easier to submit data to the state.

Changes to the data collection system may affect any item on the certificate. Any sharp discontinuity before and after a change might be an artifact of the change rather than a real difference.

Classification and coding of data on Washington State vital records follow National Center for Health Statistics (NCHS) guidelines as defined in ‘Vital Statistics Instruction Manuals.’ For details see www.cdc.gov/nchs/about/major/dvs/im.htm.

COMMENTS ON INDIVIDUAL ITEMS (alphabetically arranged)

AGE (MOTHER, FATHER)

In 1989, the mother’s and father’s birth date replaced their specified ages. As of 1989, therefore, ages are computed from the birth date and the date of delivery. A comparison of data before and after the change showed that ages calculated from birth dates are consistent with ages collected directly from the mother and that there is no substantial increase in missing data as a result of collecting the more detailed birth dates.

The father’s age is missing from a substantial number of records, mostly because the mother is unmarried and a paternity affidavit has not yet been filed.

**Please note:** Some facilities still used the old birth certificate data collection form for 1989-1990, and, therefore, did not collect mother’s DOB. As a result, mother’s DOB and father’s DOB data field may be blank for some records for those years.

ALCOHOL USE

Alcohol use data are substantially underreported on birth certificates. A national telephone survey found that only about 5% of women who drink during pregnancy report it on the birth certificate. Because of poor reporting, the 2003 birth certificate no longer collects alcohol use data.

APGAR SCORE (5 MINUTES)

In 2003, the 1-minute Apgar score was discontinued. At the beginning of the year, some hospitals were putting the 1-minute score in the place for the 5-minute score since they were used to putting in the 1-minute score first. This was corrected in the middle of the year. For 2003, users should just use the later data for 5-minute score (August-December).

ATTENDANT/CERTIFIER CLASSIFICATION

The birth certificate collects data on both the certifier (who attests that the child was born alive at the time, place, and date stated) and the attendant (who actually delivered the baby). Since mid-1985, the file has codes for the classification or title of the certifier and the attendant (e.g., MD, licensed midwife). However, the attendant’s data are given only if s/he is different from the certifier. Before mid-1985 the certifier class field was not used. To analyze complete data by attendant type, use the following guidelines:

1. Before mid-1985: Use the attendant classification.
2. Mid-1985 and after: Select the attendant classification if given. Otherwise, use the certifier classification.

Coding issues for this item are:

1. 1987: Changed the meaning of three codes (02, 04, and 05). See the Data Dictionary for details.
2. 1996: Incorrectly coded certified midwives (code 05) as nurses (code 06). The data were corrected for major facilities by using a list of certified midwives to reset the code. However, the codes may still be incorrect for some of the smaller facilities.
3. 1998: Improved the coding of midwives by using the name of the midwife to assign a code when the title was not given
4. 1999: Increase in the number of births with a hospital administrator given as the birth attendant. In these cases the hospital administrator was the certifier but the attendant classification was missing. According to the rules the hospital administrator thus becomes the birth attendant. The Center for Health Statistics is working with facilities to correct this reporting problem. As of 2003, the problem has nearly been eliminated.

BIRTH WEIGHT

Birth weight is given as grams on the data file. Many scales weigh babies in pounds and ounces, which the computer converts to grams. In 1980-91, even if the weights were reported in grams, they were converted to pounds and ounces at data entry, then reconverted to grams for data analysis. Since one ounce is equivalent to 28 grams, these converted gram weights will cluster in multiples of 28 grams and will probably not be the same as the gram weights originally reported. Starting in 1992, weights in grams were directly entered into the computer, so that values which are not multiples of 28 will be found for these years.

BODY MASS INDEX (BMI)

The Body Mass Index (BMI) is a measure of weight for height. The formula for calculating Body Mass Index is: BMI = 703.1 x (prepregnancy weight in lb / square of height in inches). For analysis, Body Mass Index is generally grouped as follows: Underweight (<18.5), Normal (18.5 – 24.9), Overweight (25.0 – 29.9), and Obese (30.0 and above).

CERTIFICATE NUMBER

In 2007, increases in the number of babies born in the state required the Center for Health Statistics (CHS) to change the certificate numbering scheme used in previous years. The range of certificate numbers for in-state births changed from 1-88,999 to 1-199,999. Since the certificate number has always been 6 characters (plus the birth year) the field size did not expand.

Public use data files do not include the actual certificate number. Instead, they have an encrypted number. In previous years these encrypted numbers have all started with ‘7’ to distinguish them from actual certificate numbers. Because of the expansion in certificate number range, CHS added more encrypted numbers to cover the new ranges. These new numbers start with ‘5’ or ‘6,’ are still distinct from actual certificate numbers, and are still 10 characters long (4-digit birth year plus 6-digit encrypted number).

CERTIFICATE TYPE (DELAYED REGISTRATION)

Births registered more than four years after the date of birth are called “Delayed Registrations” and are assigned a Type D Birth Certificate. These types of certificates record a very limited amount of information due to the delay between the date of birth and the date of registration. They are included here for completeness sake.

CITY/COUNTY/STATE OF RESIDENCE AND OCCURRENCE

*County coding:* In earlier years, the county of residence was based on reporting by the mother. As of 1997, the mother’s residence county is coded by the DOH (Department of Health) Standard Process for Matching and Geocoding, which uses a variety of matching maps and software to assign a county based on the mother’s residence address. The county assigned by the geocoding software differs from the county of residence reported by the mother for a small number of records (< 0.5% of all births). In most instances where differences are found, the geocoded county is correct and, in those instances, it is used in place of the reported county.

*City coding:*  The city of residence or occurrence is only coded if it has at least 2,500 people. Otherwise, it gets a ‘balance of county’ code ('00'), along with other small areas in the county. A city near the cutoff point may fluctuate above and below 2,500 and thus may have a separate code in some years and not in others. A count of zero births in a particular year may simply mean that the city did not have a separate code in that year.

Population estimates provided by the Washington State Office of Financial Management establish which cities meet the population criteria for separate coding. Because these estimates are published in the middle of the year, changes do not appear in the birth data file until the following year. Thus, a city which first exceeds 2,500 population in 2000 would not have a separate code until 2001.

The city of residence code is based on whether or not the mother lives within city limits. (These data are collected from the item on the birth certificate: ‘Inside city limits - yes/no’.) If she does (or if the city limits item is blank or unknown), the residence gets a distinct city code if applicable. Otherwise, the city code is set to ‘00’.

*Unknown data:* In the few instances in which the county or city of residence or occurrence is unknown, the county/city code is imputed using NCHS guidelines.

If the county is known but the city is not, use the rural portion of the county (city code ‘00’). Otherwise:

1. For Washington occurrence births, use the county and city of occurrence of the previous record.
2. For Washington residence births,
3. If the birth occurred in Washington, use the county/city of occurrence.
4. If the birth did not occur in Washington, use the largest city in the state (Seattle, code ‘1701’).

*Selecting residence or occurrence:* All files have data for both the mother’s place of residence and the place where the birth occurred. To study Washington State residents, select state of residence (st\_res) = ‘48’. Similarly, to study Washington State occurrences, select state of occurrence (st\_occ) = ‘48’.

FACILITY OF BIRTH

Two things can affect the number of births in a particular facility:

1. A change in facility code due to changes in licensing (e.g., a new name, new ownership)
2. A change in facility characteristics such as a merger with another facility or changes in facility service area

Thus, a given facility code may have several births in one year and then drop to zero the next year – or may have a large increase in births. Every year, the Center for Health Statistics examines time trends in births by facility and verifies any unusual pattern with the facility.

Midwives (particularly licensed midwives) have two series of facility codes. The ‘300’ series is for midwives delivering at the mother’s home or someone else’s private home. The ‘400’ series is for midwives who deliver at their place of business, typically a licensed birthing center. A single midwife may have two codes if s/he delivers in both places.

GESTATIONAL AGE – CALCULATED AND CLINICAL ESTIMATE

The birth certificate provides for 2 ways of determining gestational age:

1. By calculation from the menses date and the birth date
2. By a clinical estimate

*Calculated age:* The gestational age in weeks is calculated by subtracting the date of last normal menses from the birth date, dividing by 7 and truncating the result to eliminate decimal places. If the menses day is missing but the month and year are present, a value of ‘15’ is used for the day.

Prior to 2005, if the menses month and/or year were missing or the calculated gestational age was beyond a reasonable range (<18 or >45 weeks), the gestational age was estimated from the child’s birth weight.

Currently, if the gestational age cannot be calculated because of missing menses dates or the calculated age is out of range, the clinical estimate is used. If the clinical estimate is also out of range or unknown, the calculated age is unknown. This change makes the Washington State data consistent with data published by NCHS.

For 1980-88, the birth certificate did not collect the clinical estimate. For these years, only the gestational age calculated from the menses date is included in this field. In all other cases, the calculated age is unknown.

For data analysis, NCHS recommends using the clinical gestational age rather than the calculated estimate. The gestation flag field identifies which ages were calculated from the menses date and which were imputed from the clinical estimate.

*Clinical estimate:* Compared to the calculated gestational age, the clinical estimate of gestation has a very strong peak at 39 weeks (‘term’ birth). A 2015 study (National Vital Statistics Reports, Vol. 64 No. 5 June 1, 2015) found that the two gestational ages agree exactly for 62% of the births and within one week for 83%. Agreement is best for babies born near normal term (38-42 weeks). The clinical estimate exhibits lower levels of preterm and postterm births and higher levels of births at full term.

Increasing evidence of greater validity of clinical estimates compared to calculated estimates, and the national availability of clinical estimates data, have prompted NCHS to use the clinical estimate as its primary measure of gestational age beginning with 2014 data.

INFANT DEATH FLAG

The infant death flag is set to ‘X’ if the infant died by one year of age. This information comes from linking infant death records with birth certificate information. This flagging is not done for infants born in the most recent year because one year has not elapsed for all birthdates in that year. Misleading conclusions could be drawn from partial flagging of infant deaths for current year births.

The birth-infant death record linkage program has become more thorough since the mid-1980’s. Therefore, some infant deaths in the early 1980’s may not be flagged. Interpretation of the infant death data in those years should be done with extreme care.

INFANT LIVING AT TIME OF REPORT

This item was added to the birth certificate in 2003. A check against death records showed that many of the records marked ‘No’ did not have a matching death certificate. These records should have been marked ‘Y’, indicating that the infant was still alive. This problem has been corrected starting with the 2005 data. It was not possible to go back and correct the 2003 and 2004 data. Therefore, for these 2 years, the infant death flag should be used instead to study birth data for infants who have died.

MATERNAL SMOKING (YES/NO AND NUMBER OF CIGARETTES PER DAY)

This item has undergone wording and placement changes over time. Note that data may not be comparable before and after the change. Use caution in doing any trend analysis which spans the change.

1. 1984-88: Used wording 'Smoking at any time during the pregnancy' and placed in the middle section of the certificate, which the mother generally completes from a worksheet.
2. 1989: Changed wording to 'Tobacco use during pregnancy' (which could include smokeless tobacco) and relocated to the bottom of the certificate, which is generally completed by the physician. The percentage of missing data increased from 4% in 1984 to 13% in 1989, possibly as a result of this change.
3. 1992: Changed back to original wording and placement on the certificate
4. 2003: Item revised to collect average number of cigarettes per day three months before pregnancy and by trimester during pregnancy, but placement not changed.

MEDICAL AND HEALTH INFORMATION SECTION (e.g., congenital anomalies of the newborn)

This section has changed considerably over time. Because of these changes, a particular code or box number does not necessarily mean the same thing from year to year. For example, a code '6' for method of delivery means 'Version & Extractions' for 1980-83, 'Repeat C-Section' for 1984-88, 'Forceps' for 1989-2002, and does not have a definition starting in 2003. In studying trends for a particular condition, make sure to select the correct codes for each year (see Data Dictionary).

Changes in individual reporting practices and definitions and in the data collection process affect data in this area. For example:

1. In 1989, the positioning of the 'none' box for all of these items was scrambled, so they weren’t all lined up across the top of the section, as they had been previously. Thus it was harder to check 'none' straight across for all items. Possibly because of this change, the number of certificates marked 'none' decreased in 1989.
2. Before 1996, the DCT did not allow ‘unknown’ to be entered for any item in this section, so ‘none’ was often entered instead. The new DCT made it possible to enter ‘unknown,’ which may have increased the number of unknowns, at the expense of responses coded ‘none.’ Thus, the change from ‘none’ to ‘unknown’ is difficult to interpret. The data are different but it is not clear whether they are better or worse.
3. Washington State birth certificates may overestimate two conditions because of reporting practices: Rh sensitization: Hospitals may be reporting Rh incompatibility rather than Rh sensitization (which is rare). Other excessive bleeding: Hospitals may also be misinterpreting the definition of excessive bleeding. Both of these conditions are much more often reported on Washington State birth certificates, compared to US figures.
4. Starting in 1999, placenta previa decreased in frequency. Training efforts around the new birth certificate system provided clearer definitions of what should be included with the various items. Thus placenta previa may have been over reported in the past.

The Center for Health Statistics has begun providing standard definitions for the items in this section, to improve comparability between facilities

The current item ‘Risk factors in this pregnancy’ was called ‘Medical risk factors for this pregnancy’ in 1989-2002, ‘Complications during or related to pregnancy’ in 1984-88, and was two separate items, ‘Complications related to pregnancy’ and ‘Complications not related to pregnancy’ in 1980-83. Data for ‘Complications related to pregnancy’ are included in the ‘Risk factor’ field and ‘Complications not related to pregnancy’ are in a separate field (see Data Dictionary).

The item 'Congenital anomalies of child' was an open-ended question in 1980-83, which means that hospital personnel had to write in the appropriate condition. Because open-ended questions generally have less complete reporting than check-box questions, the 1980-83 data for this item may not be comparable to later data. These 1980-83 data are included on the file as four-digit ICD-9 codes in fields separate from the rest of the malformation/complication section (see Data Dictionary).

For 1992-1995 infant deaths due to congenital anomalies, the corresponding birth data were examined to see if the anomaly had been reported on the birth certificate. Reporting rates ranged from 11% for circulatory/respiratory system anomalies to 64% for anencephalus. These percentages give an estimate of how well congenital anomalies are reported on birth certificates.

 ‘Abnormal Conditions of the Newborn’ includes an item for ‘Assisted ventilation immediately following delivery.’ For 2003 and 2004 births occurring at Sacred Heart Hospital, Spokane, this item was overestimated. The box was checked if any little whiff of oxygen was given to the baby. The data have been corrected starting in 2005.

MENSES DATE

This item consists of three parts - month, day, and year. Facilities may report part of the date, most often month and year, but not day. The discussion under ‘Gestational age' compares gestational ages calculated from the menses date with clinical estimates of gestational age.

MONTH PRENATAL CARE BEGAN

This item has a high percent of unknowns. However, birth certificate prenatal care data still reliably reflect the prenatal care experience of mothers in the state. This finding came from using 1996-1998 data from a survey of mothers (PRAMS, the Pregnancy Risk Assessment Monitoring System) to compare women who had missing month prenatal care began on the birth certificate to women who had month care began given. The comparison showed no significant differences between these two.

As of 2003, this field is calculated from the menses date and the date of first prenatal care visit. Because the dates are harder to get, this item now has more unknowns than in the past. In fact, unknowns more than doubled between 2002 and 2003 (from 8.9% of births to 19.6%), even using ‘15’ if the day is missing but the other parts are known.

Since the 2003 birth certificate revision created the need to calculate the month of pregnancy when prenatal care began. The Washington State and the National Center for Health Statistics did this calculation independently, and the numbers differ considerably. Most of the difference is related to the way in which the calculation converts fractional months to whole months. The WACHS formula uses the standard way of converting fractions to whole numbers. The NCHS formula assumes that any fractional value greater than the exact month means that care began in the next month. Thus, for example, a calculated month of 2.1 would be the 2nd month by the WACHS formula and the 3rd month by the NCHS formula. Since most states have now converted to the 2003 U.S. birth certificate we have added the field, ‘pnatfed’ using the NCHS method to the data file for comparability with other revised states. However, when looking at WA State trends, the ‘pnatalmo’ field can still be used.

Please note: the initial program used to calculate month prenatal care began for 2003 data had a slight error. This program has now been corrected and the data recalculated for 2003. The main effect of this recalculation is that 568 records previously assigned to unknown month care began are now assigned to ‘no prenatal care.’

OCCUPATION (MILHAM CODES - MOTHER, FATHER)

The Milham occupation codes were developed in Washington State to manually code mother’s and father’s occupation in 1980-91. Starting in 1992 the computer assigns Milham codes based on the occupation literals where there is an exact match. Records where there is no computer match are not edited or coded, thus only about 93-97% of these records have Milham codes. These codes are not necessarily comparable to Census Bureau occupation and industry codes.

PREVIOUS PREGNANCY HISTORY (total prior pregnancies, live births now living and now dead, fetal deaths and other terminations)

These items have undergone several changes since 1980, primarily in the amount of detail (gestational age breakdowns) available for specific outcomes. Refer to the Data Dictionary to see what data are available each year.

Coding issues for this item are:

1. 1985-88: Did not collect data for the Date of Last Fetal Death. The certificate only collected Date of Last Other Pregnancy Outcome, which could either be a fetal death or an induced termination. For births where the mother had no induced terminations but reported at least one prior fetal death, use the Date of Last Other Pregnancy Outcome as the Date of Last Fetal Death.
2. 1989: Edited number of previous pregnancies so that if the sum of the detailed outcomes equals the total number of prior pregnancies specified, then any blanks in the detailed outcomes are replaced with ‘0’. This edit substantially decreased the number of unknowns caused by facilities not putting ‘0’ where applicable. Users can perform a similar edit in earlier years to reduce the number of unknowns.
3. 1999: Eliminated the total number of prior pregnancies as a separate item. It is now computed from the detail supplied. The data are now edited so that if at least one field contains an entry, then any blank field is converted to zero.

RACE/HISPANIC ORIGIN (MOTHER, FATHER, CHILD)

Race/Hispanic origin for the mother, father, and child (selected years) are collected by asking the mother for the data. In addition to the designated data, the child's race and Hispanic origin are calculated from the mother's and father's race/Hispanic origin by the following formula:

1. If either parent is Hawaiian, the child is Hawaiian. Otherwise:
2. If one parent is non-White, the child is given that race.
3. If both parents are non-White, the child is given the father’s race.
4. If one parent’s race is unknown, the child is given the known race.
5. If both parents’ races are unknown, the child’s race is unknown.

The following changes have been made to the collection and coding of race and Hispanic Origin since 1980. Note that data may not be comparable before and after the change. Use caution in doing any trend analysis which spans the change.

1. Mid-1988: Expanded race code 'C' (which previously had been used only for Mexican/Chicano) to include all Hispanics
2. 1988: Added a separate item for Hispanic Origin to the birth certificate.
3. 1990: Changed race coding in accordance with NCHS guidelines. If more than one race was given on the birth certificate, assign the first race listed as the person's race. Before 1990, the first non-white race was used. This change increased the number of births coded as “White.”
4. 1992: Expanded race codes to provide more detail on Asian Pacific American subgroups (see Data Dictionary). However, a large number of records still report just ‘Asian’ as the race. Thus, birth rates for Asian subgroups may substantially underestimate the actual rates.
5. 1992-2002: Allow the mother to designate a race/Hispanic origin for her child.
6. 1996: Modified Hispanic origin coding to comply with National Center for Health Statistics (NCHS) guidelines. Use the person’s race or place of birth to provide more specificity on country of origin. See the NCHS instruction manuals cited above for more information.
7. 1998: Added specific codes for two Hispanic categories previously classified “Other Hispanic” (code 5): “Boricua” now “Puerto Rican” (code 2), and “La Raza” now “Central or South American” (code 4).
8. 2003: Added check boxes for race. Parents may now report more than one race. Discontinued child’s designated race/Hispanic origin. See note below.

Changes to race/Hispanic Origin in 2003: Starting with 2003, the file has three race fields and two Hispanic fields for the mother and father:

 Bridged race: This is the multiple race data bridged back to a single race by NCHS. See http://www.cdc.gov/nchs/nvss/bridged\_race.htm for details on the bridging process. Bridged race data are in the old mother’s, father’s, and calculated child’s race fields. Use these data to study trends.

 Reported race/Hispanic: This is the mother’s and father’s race/Hispanic Origin exactly as reported on the Birth Filing Form. There is a separate field for each check box on the form.

 Summary race: This is the multiple race data edited by NCHS to remove duplicate entries. See http://www.cdc.gov/nchs/data/dvs/multiple\_race\_docu\_5-10-04.pdf for details (this is in .pdf format; you will need to download the free Adobe Acrobat reader to open it). The Center for Health Statistics (CHS) has done further recoding to group the races into the five basic groups and to create single fields for the mother’s and father’s races with a code for each possible race combination. These fields are at the end of the file. Use these data to study multiple races. In 2005, about 3% of mothers reported having more than one race. This percentage agrees well with the value of 3.0% based on Census data.

 Parent and child Hispanic: The check box data were edited by NCHS to take account of ‘other’ entries where codeable detail was provided. These edited data are in the old mother’s, father’s and calculated child’s Hispanic fields. The Hispanic data do not need to be bridged because only one category was selected.

WEIGHT GAINED DURING PREGNANCY, PREPREGNANCY WEIGHT, AND DELIVERY WEIGHT

In the early version of the DCT, a hospital could not directly enter weights of more than 250 lbs. They had to enter 250 lbs and put the true weight in a comment field. As this comment field was not often used, there was a cluster of births at 250 lbs. Even in the new DCT, weights of over 300 lbs still had to be entered in the comment field so there may be a clustering of weights at 300 lbs. The upper limit was changed to 350 lbs in 2003.

Before 2003, the birth certificate asked for the weight gain during pregnancy. The weight gain item had a high percent missing because many hospitals do not have this information on their data collection worksheets. The 1996 revision of the DCT allowed for negative weight gain (weight loss), which in prior years had to be coded ‘000.’

As of 2003, weight gain is no longer collected as a separate item. It is calculated from the prepregnancy weight and weight at delivery.

ZIP CODES - REPORTED AND GEOCODED

The zip code determined by geocoding addresses (geozip) was added to files starting with 1988 data. See ‘City/county/state of residence and occurrence’ for a description of the geocoding process. The geocoded zip code may differ from the reported zip code because: (1) the zip code was reported or data entered incorrectly; (2) the zip code boundary changed; or (3) the geocoding process matched the address to the wrong location.

In 1999, the reported zip code differed from the geocoded zip code for about 4% of residence death records. A manual review of these records found that about 93% of the geocoded zip codes were correct and that most of the errors were due to data entry error or reporting error. The birth data were not examined in this way but are expected to be similar.

For data analysis, use the geozip field in preference to the reported zip code when the two do not agree. Use the reported zip code when the geozip field is missing. Zip code boundaries change over time, however, so use caution when analyzing earlier data, especially before 1994. The number of addresses with differences between reported and geocoded zip codes will increase for older files because the geocoded zip field will have different boundaries.