An Evaluation of Competing Risks in Studies of Perinatal Mortality and Birth Defects

Imputation of Missing or Invalid Gestational Age

To avoid a further downward bias in gestational age-specific mortality, we imputed gestational age for infants with invalid or missing gestational age data using multiple imputation based on predictive mean matching.[5] We obtained a single imputation of gestational age using a model that included birth weight, plurality, live birth order, infant sex, location of delivery (hospital, birthing center, home, other), delivery method, maternal age, and maternal race/ethnicity.

Results of Gestational Age Assessment and Imputation

Cleaning the gestational age data led to the identification of 15,805 (0.4%) records with no plausible gestational age. Of those with no plausible gestational age, approximately half also either had a birth weight <350g (n=4,509) or were missing birth weight (n=2,751) and therefore were excluded from the analysis. Among live birth survivors and perinatal deaths, gestational age was imputed for records without a plausible gestational age but a birth weight ≥350g. Invalid or missing gestational age estimates were found for 0.2% (n=6111) of live birth survivors, 2.4% (n=539) of neonatal deaths, and 7.9% (n=1895) of fetal deaths. These records were kept if the imputed gestational age was ≥20 weeks (n=8,486). Cleaning resulted in 645,913 records being recoded from the LMP estimate to the obstetric estimate (approximately 16% of the eligible cohort). LMP estimates were retained for 84% (n=3321932) of live births surviving to one year, 80% (n=17510) of neonatal deaths, and 73% (n=17511) of fetal deaths.

Implausible estimates leading to exclusion or imputation were most common among fetal deaths and more common among infant deaths than among one-year survivors. The proportion of records recoded to the clinical estimate was similar for the one-year survivors (16.1%) and infant deaths (16.9%, n=4005) but slightly higher for fetal deaths (18.8%, n=4495). Among fetal deaths, 23,313 records had a plausible gestational age of <20 weeks and were excluded. These records are included in the fetal death file because some US states require reporting of fetal deaths <20 weeks gestational age.[6]

Chapter 2

Supplemental Methods

Missing Pregnancy Outcome and Gestational Age

Pregnancy outcome is based on maternal report from the interview for those who were interviewed and from medical records for those who were not interviewed. When pregnancy outcome was missing for those who were interviewed, outcome from medical records was substituted when available. When pregnancy outcome remained missing, the pregnancy outcome was assumed to be termination if the gestational age value after cleaning was <20 weeks.

Gestational age is based on maternal report from the interview for those who were interviewed and from medical records for those who were not interviewed. For interviewed cases with missing values of gestational age, the value abstracted from the medical record was substituted where available. Then, for both interviewed and non-interviewed cases who were still missing gestational age, values from vital records, either gestational age based on last menstrual period or based on the clinical estimate, were substituted.

Implausible Data

Birth weights for all cases were based on medical records; there were no missing values. Birth weights <100 g and birth weights <1000g among live births with a gestational age over 35 weeks were considered implausible and set to missing.

Gestational Age Cleaning

The algorithm based on an algorithm by Basso and Wilcox, [2] which compares birth weight against a gestational age and sex specific standard to create a Z score, and then considers the birth weight to be plausible if the z-score is within a set range. The range is wider (+/- 5 standard deviations) for term than preterm gestational ages (+3/-4 standard deviations when two estimates were available, +2/-3 when only one estimate was available) to account for greater inaccuracy in preterm birth weight for GA values.

To evaluate the plausibility of gestational age values for terminated cases, which often lacked a birthweight value, we created an additional algorithm. When there were two estimates of gestational age and both estimates were both <20 weeks and were ≤2 weeks apart, we kept the interview value was preferentially kept in the new gestational age variable; if the estimates were more than two weeks apart, we preferentially kept the interview value in the new variable. When there was a single gestational age estimate, we kept the value of that estimate in a new gestational age variable if the value was <25 weeks. When both gestational age variables were missing, the new variable was set to missing.

For remaining cases, when birth weight was missing and there were two available gestational age estimates, we did the following: if both estimates agreed exactly, the value was retained. If the two estimates differed by 2 weeks or less, we retained the value from the medical record; if they differed by more than two weeks, the value in the new variable was set to missing. Then, for live born cases with at least one gestational age estimate that had not met a previous recoding rule, if at least one gestational age value was at least 15 weeks, the value was retained. For stillborn cases with at least at least one gestational age estimate the value was retained if at least one estimate was at least 17 weeks, otherwise it was set to missing.

For records with a birth weight, we created sex and gestational age-specific birth weight z-scores for the medical record value and interview value using standards from Talge et al 2014. [4] Z-scores were calculated by subtracting the mean sex and gestational age-specific birth weight from the recorded birth weight and dividing by the sex and gestational age-specific standard deviation.

Gestational ages over 44 weeks were considered implausible, therefore where both gestational age estimates were > 44 weeks, we set the new variable to missing. Then, for cases with a birth weight and at least one gestational age estimate, the z-score for each gestational age estimate was examined to see if it was plausible for the stated gestational age according to the z-score cut-points set by Basso and Wilcox as described below.

When two gestational age records were available, we did the following: when the two estimates were within two weeks of each other and the medical record value was greater than 37 weeks, we first checked the medical record value estimate and set the new gestational age variable equal to its value if it was in the acceptable range (z-score of -5 to 5). If the medical record value estimate was not in the acceptable range, we then checked the z-score for the interview value. If neither estimate was in the acceptable range, the gestational age for the record was set to missing.

Where the medical record value and interview value estimates were not within two-weeks of each other, we repeated this procedure but checked the interview value estimate first, moving to check the medical record value estimate if the interview value estimate was not in the acceptable range. For records with an interview value <37 weeks, and for which the two estimates were within two weeks of each other, we used the same procedure comparing to a smaller acceptable range (z-score of -4 to 3), moving to check the medical record estimate second if the interview estimate was not in the acceptable range. If neither estimate was in the acceptable range, the new gestational age variable was set to missing.

Where only one gestational age estimate was available the estimate was checked against the relevant criteria (-5 to 5 for term births and -4 to 3 for preterm births) and retained where the criteria was met or gestational age variable was set to missing if the criteria was not met. We followed the same procedure for records in which one gestational age estimate was <37 weeks while the other was ≥37 weeks.

For stillbirths without an identified gestational age estimate after this procedure, if both estimates were within two weeks of each other, then we retained the value for the medical record value, otherwise the new gestational age variable remained set to missing. For stillbirths that were part of a multiple gestation pregnancy, to account for possible delayed delivery, we retained the recorded gestational age value if at least one z-scores was negative, regardless of the value of the z-score.

Then, for those whose gestational age was missing or <20 weeks following the completion of the algorithm, we further substituted gestational age values from medical records or the fetal death certificate if at least one of these estimates was ≥20 weeks. We used the same procedure to fill in gestational age data for terminated cases whose gestational age was still missing and for whom at least one gestational age value was <26 weeks (when most terminations of cases with available gestational age data were identified in our data). Two cases were explicitly recoded based on data from birth certificates showing errors in the medical record value or birth weight.

Then, we created a category for whether or not the cleaned gestational age value was above or below 20 weeks gestation.

For those missing a cleaned gestational age value and two gestational age values, when both were ≥20, the value was considered to be <20 weeks. For those where at least one gestational age variable had a value of <20 weeks or who only had one or no gestational age estimates, if birthweight was <500g, the value was considered to be <20 weeks; if birthweight was >=500g, the value was considered to be >20 weeks. If a record did not meet any of the above criteria and the birthweight was missing, the value of the indicator was set to missing.

National Birth Defects Prevention Study Eligible Birth Defects

- Anencephaly and craniorachischisis
- Spina bifida
- Encephalocele, cranial meningocele, encephalomyelocele
- Holoprosencephaly
- Hydrocephalus
- Dandy-Walker malformation
- Anophthalmia/microphthalmia
- Cataracts, glaucoma, and related eye defects
- Anotia/microtia
- Conotruncal heart defects
- Single ventricle
- Septal heart defects (atrial septal defects, ventricular septal defects)
- Atrioventricular septal heart defects
- Ebstein malformation
- Obstructive heart defects (right and left ventricular outflow defects)
- Anomalous pulmonary venous return
- Heterotaxia
- Choanal atresia
- Cleft lip with and without palate

- Cleft palate
- Esophageal atresia with and without tracheoesophageal fistula
- Intestinal atresia/stenosis
- Biliary atresia
- Hypospadias (third or second degree)
- Bilateral renal agenesis or hypoplasia
- Limb deficiency, intercalary
- Limb deficiency, longitudinal
- Limb deficiency, transverse
- Limb deficiency, not elsewhere classified
- Craniosynostosis
- Diaphragmatic hernia
- Sacral agenesis
- Omphalocele
- Extrophy, bladder
- Extrophy, cloacal
- Gastroschisis
- Amniotic band syndrome

For further details, see Yoon et al 2001. [7]

References

1 World Health Organization. The WHO application of ICD-10 to deaths during the perinatal period: ICD Perinatal Mortality (ICD-PM). Geneva, 2015.

2 Basso O, Wilcox A. Mortality risk among preterm babies: immaturity versus underlying pathology. *Epidemiology* 2010; **21**: 521–7.

3 Martin JA, Wilson EC, Osterman MJK, Saadi EW, Sutton SR, Hamilton BE. Assessing the quality of medical and health data from the 2003 birth certificate revision: results from two states. *Natl Vital Stat Rep* 2013; **62**: 1–19.

4 Talge NM, Mudd LM, Sikorskii A, Basso O. United States birth weight reference corrected for implausible gestational age estimates. *Pediatrics* 2014; **133**: 844–53.

5 Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica* 2014; **68**: 61–90.

6 Martin JA, Hoyert DL. The national fetal death file. *Seminars in Perinatology* 2002; **26**: 3–11.

1 World Health Organization. The WHO application of ICD-10 to deaths during the perinatal period: ICD Perinatal Mortality (ICD-PM). Geneva, 2015.

2 Basso O, Wilcox A. Mortality risk among preterm babies: immaturity versus underlying pathology. *Epidemiology* 2010; **21**: 521–7.

3 Martin JA, Wilson EC, Osterman MJK, Saadi EW, Sutton SR, Hamilton BE. Assessing the quality of medical and health data from the 2003 birth certificate revision: results from two states. *Natl Vital Stat Rep* 2013; **62**: 1–19.

4 Talge NM, Mudd LM, Sikorskii A, Basso O. United States birth weight reference corrected for implausible gestational age estimates. *Pediatrics* 2014; **133**: 844–53.

5 Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica* 2014; **68**: 61–90.

6 Martin JA, Hoyert DL. The national fetal death file. *Seminars in Perinatology* 2002; **26**: 3–11.

7 Yoon PW, Rasmussen SA, Lynberg MC, *et al.* The National Birth Defects Prevention Study. *Public Health Rep* 2001; **116**: 32–40.