OBSTETRICS

Spontaneous and indicated preterm birth subtypes: interobserver agreement and accuracy of classification

Molly J. Stout, MD, MSCI; Rachell Busam, RN; George A. Macones, MD, MSCE;
Methodius G. Tuuli, MD, MPH

OBJECTIVE: The purpose of this study was to estimate interobserver variability and correct classification of preterm birth into spontaneous and indicated subtypes.

STUDY DESIGN: This was a cross-sectional study in which a trained obstetric nurse, maternal fetal medicine (MFM) fellow, and MFM faculty member independently reviewed discharge summaries and full medical records to classify preterm birth into "spontaneous" and "indicated" subtypes. Consensus classification was obtained with a senior faculty member and was considered the correct classification. Proportions of correct classification by both discharge summary and full medical record review and by level of reviewer were compared with the use of the χ^2 test. Interobserver variability was estimated with the use of Fleiss' kappa.

RESULTS: Of 132 preterm births, 58.8% were spontaneous. Interrater agreement for classification of preterm birth subtype based on the full medical record review was substantial (0.79; 95% confidence interval, 0.76—0.80). Interrater agreement was slightly less, based on discharge summary classification alone (Kappa, 0.73; 95% confidence interval, 0.71—0.79) compared with a full medical record review, but this difference was not significant ($P = .3$). Correct classifications for research nurse, MFM fellow, and MFM faculty member were 85%, 95%, and 93%, respectively, for the full medical records and 85%, 93%, and 92%, respectively, for the discharge summaries alone. There was no significant improvement in correct classification based on full medical record review compared with discharge summary alone for any level of reviewer ($P > .6$).

CONCLUSION: There is substantial, but imperfect, agreement between reviewers for classification of preterm birth into spontaneous and indicated subtypes. Incorrect classification may occur 5-15% of the time, even with experienced research personnel. Discharge summaries that are populated with pertinent clinical data may streamline accuracy for research efficiency.

Key words: interobserver agreement, preterm birth, research

Preterm birth (PTB) at <37 weeks of gestation continues to be a vexing problem for obstetricians and pediatricians and is the major determinant of perinatal morbidity and death.[1] Despite agreement that PTB remains a major public health problem, lack of progress in gaining a clear understanding of pathophysiologic pathways and strategies for PTB prevention and treatment may be explained in part by the multiple and highly diverse phenotypes of PTB.

Traditionally, PTB because of preterm labor with cervical dilation or preterm rupture of membranes is classified as 'spontaneous.' Labor that is induced or in which the infant is delivered by cesarean section for maternal or fetal illness is classified as 'indicated' PTB. This splitting of PTB phenotypes is one attempt to separate distinct pathophysiologic pathways and patients who may benefit from different prediction, prevention, and treatment strategies.

A group of expert PTB researchers recently met to discuss considerations and challenges of classification of the PTB syndrome.[2-4] The consensus from this panel is that accurate classification of PTB subtypes is important both for epidemiologic surveillance and for research progress. However, definitions of spontaneous and indicated PTB may vary by individual interpretation of obstetric providers and research personnel. Incorrect classifications of PTB subtypes may lead to clinical heterogeneity and ultimately biased findings in observational and interventional studies. We hypothesize that there is

significant interobserver variability in the designation of PTB as spontaneous or indicated and that the accuracy of classification depends on the completeness of the medical record that is used for review. We aimed to estimate the interobserver agreement and accuracy among reviewers with different experience levels for the designation of PTB subtypes based on discharge summaries and complete medical record review.

MATERIALS AND METHODS

We performed a cross-sectional study in which a trained research nurse (registered nurse with clinical background in labor and delivery nursing whose full-time job is management and coordination of research studies), a maternal fetal medicine fellow, and a maternal fetal medicine faculty member independently reviewed randomly selected medical records of patients who delivered at <37 weeks of gestation. These 3 reviewers were chosen at random (and in the case of fellow and faculty reviewers from those with available research time for this project). PTB patients who enrolled into the Women's and Infants Health Specimen Consortium, an ongoing tissue and clinical data collection core at Washington University in St. Louis. MO, were included. The Washington University School of Medicine Human Subjects Review Board approved this study.

Each of the 3 reviewers (research nurse, maternal fetal medicine [MFM] fellow, MFM faculty member) first reviewed the discharge summary alone and classified the PTB as spontaneous or indicated. The entire study patient list was classified with the use of only the discharge summary before moving on to the complete medical record. Subsequently, the entire medical record was reviewed, and the PTB was again classified as spontaneous or indicated. The classification of the discharge summary was not changed based on the findings of the full medical record.

At our institution, an electronic discharge summary is populated automatically based on the information contained in the physician note completed at the time of delivery and includes fields such as: gestational age, labor, vaginal delivery, cesarean section, anesthesia, delivery date and time, Apgar score, and abnormal conditions. If a dictated discharge summary was present, it was eligible for review for the discharge summary classification. All documents in the medical record were eligible for review for the full-chart review classification and typically included: admission history and physical progress notes, delivery note, dictated operative note, pathologic report, and discharge summaries.

After individual classification based on the discharge summary and the full medical record was completed, the 3 reviewers met jointly with a senior MFM faculty member for a consensus classification. The 4 reviewers (research nurse, fellow, faculty member, senior faculty member) reviewed the complete medical record and decided by consensus whether the delivery should be classified as spontaneous or indicated. The consensus classification was considered the "correct" classification for statistical analyses.

Interobserver agreement was estimated separately based on the discharge summary and then the entire medical record for each level of reviewer. Fleiss' kappa was used to estimate interobserver agreement. Fleiss' kappa is a statistical measure for the assessment of the reliability of agreement between ≥2 raters when classifying categoric items and measures the degree of agreement in classification over that which would be expected by chance. Fleiss kappa ranges from 0—1.[5]

Accuracy of classification was performed based on discharge summary alone and full medical record review for each level of reviewer. A classification was considered correct if it agreed with the consensus classification. The proportion of correct classification for each reviewer was compared with the use of the χ^2 test.

We estimated the sample size using the method developed by Reichenheim[6] for Fleiss' kappa sample size based on the desired precision. Assuming a kappa of 0.8 (substantial agreement) and proportions of PTB classified as spontaneous by any 2 reviewers as 0.5 and 0.6, a total of 132 PTB records needed to be reviewed for an absolute precision of 0.1 for kappa.

Stata software (version 12.0; Stata Corporation, College Station, TX) was used to perform all analyses.

RESULTS

A total of 132 randomly selected PTB medical records were reviewed. Based on consensus review, 58.8% were spontaneous PTB, and 41.2% were indicated PTB. Overall interrater agreement for PTB classification based on the full medical record was 0.79 (95% confidence interval, 0.76—0.80) consistent with substantial agreement.[5] The interrater agreement was slightly less based on review of the discharge summary alone 0.73 (95% confidence interval, 0.71—0.79); however, this difference was not statistically significant ($P = .3$).

Correct classifications for research nurse, MFM fellow, and MFM faculty member were 86%, 95%, and 93%, respectively, for the full medical record review (Table). Correct classification based on the discharge summary review alone was similar and is shown in the Table. When percent of correct classification was compared, with the MFM faculty member as the referent, the research nurse classification was less accurate for both discharge summary and medical record review; however, this difference was not statistically significant. There was no significant improvement in correct classification based on the medical record review compared with the discharge summary alone for any level of reviewer. Of the 33 records for which ≥1 reviewers disagreed, 19 records (57%) were diagnoses of premature rupture of membranes with subsequent complications such as abruption or chorioamnionitis.

COMMENT

The findings of this study suggest that there is substantial, but imperfect, agreement between reviewers for classification of PTB into spontaneous and indicated subgroups. Incorrect classification of PTB subtypes occurred 5-15% of the time, depending on the reviewer. There was no significant improvement in correct classification when the full medical record