Published in final edited form as:

Matern Child Health J. 2012 August ; 16(6): 1241–1246. doi:10.1007/s10995-011-0882-x.

Is the accuracy of prior preterm birth history biased by delivery characteristics?

David N Hackney, MD, MS^a, Danielle Durie, MD, MPH^a, Ann M Dozier, RN, PhD^b, Barbara J Suter, RN, MPH^b, and J Christopher Glantz, MD, MPH^a

^aDivision of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, University of Rochester School of Medicine, Rochester, NY

^bDepartment of Community and Preventative Medicine, University of Rochester School of Medicine, Rochester, NY

Abstract

Objectives—To assess the sensitivity of birth certificates to preterm birth history and determine whether omissions are randomly or systemically biased.

Methods—Subjects who experienced a preterm birth followed by a subsequent pregnancy were identified in a regional database. The variable “previous preterm birth” was abstracted from birth certificates of the subsequent pregnancy. Clinical characteristics were compared between subjects who were correctly versus incorrectly coded.

Results—713 subjects were identified, of whom 65.5% were correctly coded in their subsequent pregnancy. Compared to correctly coded patients, patients who were not correctly identified tended to have late and non-recurrent preterm births or deliveries that were secondary to maternal or fetal indications. A recurrence of preterm birth in the subsequent pregnancy was also associated with correct coding.

Conclusions—The overall sensitivity of birth certificates to preterm birth history is suboptimal. Omissions are not random, and are associated with obstetrical characteristics from both the current and prior deliveries. As a consequence, resulting associations may be flawed.

Keywords

Birth certificate; preterm birth; prior preterm birth; reporting bias

Introduction

In 2003 the United States Centers for Disease Control introduced the new Standard Certificate for Live Birth [1], which has since been adopted by 22 states. Among the many changes in the updated certificate was the addition of “previous preterm birth”. Prior preterm birth is an important risk factor for preterm delivery in a subsequent pregnancy [2], and

Corresponding author: David N Hackney MD, MS, Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, University of Rochester Medical Center, 601 Elmwood Ave, Rochester, NY 14642, Phone: (585)-275-7480, Fax: 585-256-1416, david_hackney@URMC.rochester.edu.

many attempts at preterm birth prevention, such as 17-alpha-hydroxyprogesterone caproate [3] and cervical cerclage [4], are focused on these patients. Epidemiologically, preterm birth history can be used to assess the degree to which an obstetric population is at risk, and thus this birth certificate variable has been utilized as a confounder in regression analysis [5].

Research that utilizes birth certificate data has both advantages and disadvantages. For example, birth certificate data can be used to construct datasets that are much larger than those that could be developed at a single institution and are thus useful in the evaluation of uncommon outcomes. These advantages, however, are balanced against questions of birth certificate data accuracy, and in fact many variables have demonstrated low sensitivity or specificity when compared to “gold standards” of maternal or infant hospital charts. For example, when Reichman et al [6] compared the birth certificate report of “Previous preterm or small for gestational age infant” to the maternal delivery records from 1989–1992, they found a specificity of 99.2% but a sensitivity of only 10.7%. Overtime, measured sensitivity of preterm birth history on the birth certificates has improved; however, this improvement is still suboptimal. A high rate of omissions, however, may not ultimately be important if the omissions occur randomly. Random omissions will balance themselves out in a sufficiently large dataset, and a large sample size may compensate for the loss of statistical power from random omissions. Omissions that are biased according to clinical characteristics, however, will produce systematic errors that cannot be corrected by increasing sample size and thus lead to inappropriate results and conclusions [7].

Intuitively, several factors could potentially bias the correct determination of a patient’s preterm birth history. Deliveries at a very early gestational age, for example, may be recalled more accurately by patients and providers than prior “late” preterm births. Likewise, patients who experience a recurrent preterm birth may be more likely to report their prior history at the subsequent delivery than women who subsequently deliver at term. For example, Adams et al demonstrated that the sensitivity of preterm birth history increased when the subsequent delivery was of a low birthweight [8], and Reichman et al demonstrated an increased sensitivity for prior preterm birth when the subsequent pregnancy was also preterm [9]. The clinical circumstances surrounding the prior preterm birth may also play a role. For example, prior preterm deliveries secondary to preterm labor or membrane rupture may be correctly identified more readily than those that were intentional deliveries secondary to maternal or fetal indications. Thus the purpose of this study was to evaluate both the sensitivity of the variable “previous preterm birth” from birth certificates and assess omissions for sources of bias according to clinical characteristics and outcomes in prior and current pregnancies.

Materials and Methods

The Finger Lakes region data set comprises a nine-county region within Upstate New York that includes 13 hospitals (two level III, one level II, and 10 level I), for a total of approximately 14,500 deliveries per year. Birth-certificate coders at each hospital obtain birth certificate data using standardized definitions, largely in check-box format. Information is primarily drawn from the mother’s hospital and prenatal records. This study used hospital provided birth certificate data from January 2004 to December 2009. The study was approved by the University of Rochester Research Subjects Review Board. Definitions of

clinical variables included in the regional birth certificates are available online (10), and their sensitivity and specificity of these variables in our regional database have been previously published [11].

The database cohort was queried for all subjects delivering prematurely from 20+0/7 to 36+6/7 weeks estimated gestational age (EGA). Of note, the available gestational ages in the database are rounded down to the nearest week. Once identified, the database was re-queried for additional pregnancies from the same individual by searching with a maternal identifier utilized by the database. The validity of linked pregnancies was assessed by cross referencing the maternal name and date of birth, as well as screening delivery intervals for biologic incompatibilities. Specifically, if the interval between the two deliveries was less than the reported gestational age of the second child, then this would be biologically impossible and thus implies an error in the linkage. Subjects were included in the study if they had had one preterm delivery that was a live birth, followed by a subsequent delivery of a live birth at any gestational age of at least 20 0/7 weeks. The preterm birth did not necessarily need to be the subject's first pregnancy, and is henceforth referred to as the "index" delivery. If a subject had multiple previous pregnancies within the time period under analysis, the earliest delivery that was preterm was designated the "index" delivery, and the delivery that immediately followed was designated the "subsequent" delivery. Only two pregnancies were evaluated per subject. For all included subjects, the EGA at both deliveries and time interval between deliveries were abstracted from the birth certificates.

The variable "previous preterm birth" from the subsequent pregnancy was collected from the birth certificates of all subjects. According to the coder guidelines [11] this variable was to be positive if the physician diagnosed a "History of pregnancy(ies) terminating in a live birth of less than 37 completed weeks of gestation." Thus this variable should have been coded "yes" for all subjects by the fact that they had been included in the study. The "gold standard" in this regards was the gestational age of the "index" preterm delivery being less than 37 weeks. The gestational age was abstracted from the medical record by the trained coder according to that which had been assigned by the attending provider. Though no birth certificate variables have perfect accuracy, the accuracy of the gestational age in birth certificates is generally considered to be sufficient [6]. For this study the accuracy of this coding was additionally evaluated in one of two ways. For subjects who delivered at the University of Rochester Medical Center, the gestational age was confirmed through direct chart review. For subjects for whom medical records were not available for direct review, the gestational age in weeks from the birth certificate was cross referenced against the calculated gestational age from the reported LMP. If a discrepancy in the delivery being term or preterm existed which could not be resolved with record review, the subject was excluded from the study.

Because many variables of potential interest are not present on US birth certificates, a subgroup analysis was performed among subjects who delivered at our institution (University of Rochester Medical Center), for whom the delivery records were obtained and additional variables were abstracted. From the record review it was determined whether the index preterm birth had been "spontaneous" (secondary to either preterm labor or preterm premature rupture of membranes) or "indicated" (secondary to a fetal or maternal

complication). For deliveries that occurred before 24 weeks, we determined whether the delivery was secondary to an elective, indicated, or otherwise-intentional pre-viable labor induction versus a spontaneous delivery.

The overall sensitivity of the “previous preterm birth” designation in birth certificates was calculated. Subjects were then divided into two groups according to the accuracy of their preterm birth history coding. Differences in delivery characteristics in both pregnancies were compared between groups with the 2 sided t-test, Mann-Whitney U, or Fisher’s Exact test where appropriate. EGA at delivery for both pregnancies were additionally dicotomized into those that were “early” (less than 34 weeks EGA) versus “late” preterm deliveries and differences between accuracy groups were evaluated with Fisher’s Exact test. For the subgroup that delivered at our institution and underwent record review, differences in preterm birth etiology, multifetal gestation, and pre-viable labor induction were evaluated through Fisher’s Exact test. Logistic regression was performed with preterm birth history accuracy as the dependent variable, and variables that were significant in bivariate analysis compromising the independent variables. The adequacy of the final model was accessed through pseudo-R² and the Hosmer-Lemeshow goodness of fit test. A p-value <0.05 was considered significant. All analyses were performed on Stata 11 (College Station, Texas, USA).

Results

Between January 2004 to December 2009 there were 7,457 preterm births within the Finger Lakes region of Upstate New York, of whom 751 (10.1%) had a subsequent delivery within the same region in that time period for which records were available. 38 subjects were excluded secondary to a dating discrepancy in the birth certificate records of the index pregnancy which could not be resolved, for a final study population of 713 subjects. Demographic and clinical characteristics of the cohort are provided in Table 1. The variable “previous preterm birth” was abstracted from all subsequent deliveries. Overall, 467 subjects were correctly coded in their subsequent pregnancy, for a sensitivity of 65.5%. Clinical and delivery characteristics of subjects who were or were not correctly coded are provided in Table 2. Several significant differences in delivery characteristics existed between the two groups, such that subjects who were correctly identified as having had a prior preterm birth were more likely to have originally delivered at an earlier gestational age and were more likely to have had a recurrent preterm birth in their subsequent pregnancy. No significant differences existed with regards to the time interval between the pregnancies. The presence of a “late” preterm birth in the index delivery, defined as 34 0/7 to 36 6/7 weeks, was significantly associated with the preterm delivery having not been correctly identified in the subsequent delivery. The percentage of correctly coded subjects at different gestational ages in the index deliveries are presented in Figure 1.

A subgroup of 451 subjects delivered at the University of Rochester Medical Center and thus had records that were available for review. Delivery characteristics for the index preterm deliveries are presented in Table 3. Subjects were more likely to be subsequently identified as having had a prior preterm birth if their original delivery was “spontaneous,” as opposed to having been secondary to a maternal or fetal indication. There was no association between

coding accuracy and either multifetal gestation or pre-viable labor inductions, although only 7 subjects were identified with pre-viable inductions. The results of the logistic regression of preterm birth coding accuracy are presented in Table 4. All individual characteristics remained significant in multivariate regression. While the overall pseudo- R^2 for the resultant model was low (6.5%), there was no evidence of lack of fit (Hosmer-Lemeshow Goodness-of-fit test $p=0.31$).

Comment

Approximately one third of subjects with a known preterm birth were not identified as such in the birth certificates from their subsequent deliveries. Of concern is that the omissions were not completely random, but instead associated with delivery outcomes from both pregnancies. Patients were more likely to be correctly identified as having had a previous preterm birth if their index delivery was at an earlier gestational age, was spontaneous, and if they subsequently experienced a recurrent preterm birth. Thus, based on the experience of a region in Upstate New York, a degree of caution should be employed in the use and interpretation of this variable in epidemiological research.

Although the variable “previous preterm birth” was added to the Standard US Birth Certificate in 2003, it was predated by the similar variable “previous preterm/small for gestational age infant” and individual states had also previously recorded preterm birth history on their own certificates. The overall sensitivities of these similar fields had been evaluated in the past [6,8]. Of note, though the current sensitivity of 67% may be disappointing and warrants continued efforts at improvement, it represents a substantial improvement over prior reported birth certificate cohorts. For example, Reichman et al [6] demonstrated a sensitivity of only 10.7% for the variable “previous preterm/small for gestational age infant” from New Jersey birth certificates from 1989–92, and Adams demonstrated a sensitivity of only 28.9% from a linked database of Georgia birth certificates from 1980 to 1988 [8]. The associations between preterm birth coding accuracy and gestational age or other characteristics of the prior preterm birth have not previously been evaluated, although Adams et al demonstrated that the sensitivity of preterm birth history increased when the subsequent delivery was of a low birthweight [7] and Reichman et al [9] demonstrated an increased sensitivity for prior preterm birth when the subsequent pregnancy was also preterm. Of note, population differences may have existed between these study populations and ours, and thus the results may not be directly comparable.

A strength of this study is the use of a maternally linked database and, for those subjects who had delivered at our institution, examination of delivery records from their index preterm delivery. In general, most studies of birth certificate sensitivity use records from the current pregnancy as the “gold standard.” [6] For obstetrical history, however, a linked database [12] or records from the prior deliveries are more appropriate, because they may avoid inaccuracies in subsequent records. A similar problem has been demonstrated with regards to the route of delivery and subsequent identification of subjects with prior cesarean sections [13].

This study has several important limitations. Although birth certificates were evaluated across a multi-county geographic region which included many different labor and delivery units, it is possible that different results could be obtained in other parts of the country and thus ideally these results should be replicated elsewhere. Along these same lines, the results from the subgroup of subjects who underwent chart review should be interpreted with additional caution, because these represented a subpopulation within the larger cohort and arose from a single institution. Additionally, we were unable to evaluate subjects who had experienced a prior preterm delivery in the context of an intrauterine fetal demise, since such subjects would not have had an initial certificate of “live” birth in their index delivery and thus would not have been identified in our database. The impact of prior fetal demise on subsequent preterm birth history is not known, and thus it is unclear what the impact of this would be on our results. We also did not evaluate the impact of having had more than one prior preterm birth before the “subsequent” delivery, since the identification of these subjects would have required use of the “prior preterm birth” variable from their “index” delivery, which was also the variable that was under study. Because all subjects had had at least one prior preterm birth, however, data from additional pregnancies would not have changed the correctness of the prior preterm birth variable from the subsequent delivery.

Our cohort also consisted of subjects with a confirmed prior preterm birth followed by a subsequent pregnancy. Thus we were only able to evaluate the sensitivity of preterm birth history, as opposed to the specificity or positive predictive value, as this dataset did not contain false positives. The determination of false positives would have required that the absence of a prior preterm birth could be determined with absolute certainty, which presents significant logistical difficulties since one does not always have access to all prior delivery records for a given subject, especially if they had had prior deliveries outside of the geographic catchment area. In the absence of complete records, one could instead attempt to utilize the recorded preterm birth history from the earliest available pregnancy, however, such a technique would have utilized the very variable that was under study.

Additionally, the gestational age differences from the “index” delivery between subjects who were correctly or incorrectly coded, though statistically significant, is not necessarily large. Thus the magnitude of the bias may not completely exclude the use of this variable in epidemiological research. One potentially could continue to utilize the variable with the understanding that it represents a subgroup of individuals whose prior preterm births were more likely spontaneous deliveries at earlier gestational ages. The fact that the coding accuracy was associated with differences in outcomes in the subsequent delivery, however, raises greater concern. If patients with an adverse outcome are more likely to be correctly coded, then the magnitude of apparent associations would be artificially inflated. The possibility that coding accuracy varies according to obstetrical outcomes is additionally an important avenue of future exploration.

Acknowledgments

The authors would like to acknowledge the contributions of Joseph Duckett, who provided informatics support.

Support: Women’s Reproductive Health Research K-12: HD001332-09

References

1. Menacker F, Martin J. Expanded health data from the new birth certificate, 2005. *Natl Vital Stat Rep.* 2008; 56:1–24.
2. Iams J, Berghella V. Care for women with prior preterm birth. *Am J Obstet Gynecol.* 2010; 203:89–100. [PubMed: 20417491]
3. Meis P, Klebanoff M, Thom E, et al. Prevention of recurrent preterm birth by 17 alpha-hydroxyprogesterone caproate. *N Eng J Med.* 2003; 348:2379–2385.
4. Berghella V, Odibo A, To M, et al. Cerclage for Short Cervix on Ultrasonography Meta-Analysis of Trials Using Individual Patient-Level Data. *Obstet Gynecol.* 2005; 106:181–189. [PubMed: 15994635]
5. Donovan E, Besl J, Paulson J, et al. Infant death among Ohio resident infants born at 32 to 41 weeks of gestation. *Am J Obstet Gynecol.* 2010; 203:58 e51–55. [PubMed: 20417495]
6. Reichman N, Hade E. Validation of birth certificate data: a study of women in New Jersey's HealthStart program. *Ann Epidemiol.* 2001; 11:186–193. [PubMed: 11248582]
7. DiGiuseppe DL, Aron DC, Ranbom L, et al. Reliability of birth certificate data: a multi-hospital comparison to medical records information. *Matern Child Health J.* 2002; 6:169–179. [PubMed: 12236664]
8. Adams M. Validity of Birth Certificate Data for the Outcome of the Previous Pregnancy, Georgia, 1980–1995. *Am J Epidemiol.* 2001; 154:883–888. [PubMed: 11700240]
9. Riechman NE, Schwartz-Soicher O. Accuracy of birth certificate data by risk factors and outcomes: analysis of data from New Jersey. 2007; 197:e1–32.e8.
10. Guideline for the New York State Certificate of Live Birth and Quality Improvement. http://www.urmc.rochester.edu/flrpp/data-sharing/documents/Guidelines_09.pdf
11. Roohan PJ, Josberger RE, Acar J, et al. Validation of birth certificate data in New York state. *J Comm Health.* 2003; 28:335–46.
12. Petrini J, Callaghan W, Klebanoff M, et al. Estimated effect of 17 alpha-hydroxyprogesterone caproate on preterm birth in the United States. *Obstet Gynecol.* 2005; 105:267–272. [PubMed: 15684150]
13. Green D, Moore J, Adams M, et al. Are we underestimating rates of vaginal birth after previous cesarean birth? The validity of delivery methods from birth certificates. *Am J Epidemiol.* 1998; 147:581–586. [PubMed: 9521185]

Figure 1.
Percentage of subjects correctly coded as having had a prior preterm birth, grouped according to gestational age at the time of their index preterm delivery.

Table 1

Demographic and clinical characteristics of included subjects (n=713)

Mean maternal age at index delivery (SD)	25.3 years (5.4)
%Tobacco use	30.3
%Nulliparous at index preterm birth:	48.1
% African American:	25.4
% Hispanic	10.9
% Asian	6.2
% Native American	1.1
% Post-high school education	46.7

Table 2

Differences in delivery characteristics between subjects with correct versus incorrect preterm birth history coding in their subsequent pregnancy

	Preterm Birth History		
	Coded Correctly (n=504)	Coded Incorrectly (n=247)	p
Mean gestational age (wks) in “index” delivery	31.9	33.8	<0.0001
Late preterm birth* in “index” pregnancy (%)	45.8	69.5	<0.0001
Mean gestational age (wks) in “subsequent” pregnancy	36.2	37.3	0.0001
% recurrent preterm birth in “subsequent” pregnancy (%)	41.8	28.5	<0.0001
Mean delivery-to- delivery interval (months)	26.0	26.1	0.92
Multifetal gestation in “index” pregnancy (%)	8.6	8	0.87

*“Late” preterm birth defined as 34 0/7 to 36 6/7 weeks estimated gestational age. Two-sided Student t-test and Fisher’s Exact test are employed for continuous and binary variables, respectively.

Table 3

Differences in delivery characteristics within the subgroup of subjects delivering at the University of Rochester Medical Center with correct versus incorrect preterm birth history coding.

	Preterm Birth History		
	“Correct” (n=241)	“Incorrect” (n=210)	p-value
Spontaneous PTB in index pregnancy	74.2%	61.8%	0.005
Pre-viable labor induction *	1.2%	1.8%	0.42

Fisher’s Exact test was employed for binary variables. The remaining deliveries that were neither spontaneous nor pregnancy terminations were indicated preterm births.

PTB=Preterm birth. “Correct” = subject’s preterm birth history was coded correctly in their subsequent delivery. “Incorrect” = subject was not correctly identified as having had a prior preterm birth in their subsequent pregnancy.

* Elective, therapeutic, or otherwise-intentional labor induction from 20 0/7 to 23 6/7 weeks gestation in index pregnancy

Table 4

Logistic regression of preterm birth accuracy and delivery characteristics among subjects delivering at University of Rochester Medical Center (n=451)

	aOR	95% CI	p-value
SPTB	2.07	1.40–3.04	<0.0001
Late index PTB *	0.43	0.30–0.63	<0.0001
Recurrent PTB in subsequent delivery	1.79	1.21–2.64	0.003

aOR = adjusted Odds Ratio of being correctly coded as having had a previous preterm birth. A larger value is associated with greater odds of being correctly coded in the subsequent pregnancy. The regression model included the variables listed in the table. SPTB = spontaneous (as opposed to indicated) preterm birth in the “index” pregnancy. PTB = preterm birth.

* “Late” preterm birth defined as 34 0/7 to 36 6/7 weeks estimated gestational age.