PHARMACOEPIDEMIOLOGY AND DRUG SAFETY (2012)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3284

ORIGINAL REPORT

Algorithms to estimate the beginning of pregnancy in administrative databases†

Andrea V. Margulis[1,2]*, Soko Setoguchi[1,2,3], Murray A. Mittleman[1,4], Robert J. Glynn[1,2], Colin R. Dormuth[5] and Sonia Hernández-Díaz[1]

[1] *Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*
[2] *Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Boston, MA, USA*
[3] *Duke Clinical Research Institute and Department of Medicine, Duke University School of Medicine, NC, USA*
[4] *Cardiovascular Epidemiology Research Unit, Beth Israel Deaconess Medical Center, Boston, MA, USA*
[5] *Department of Anesthesiology, Pharmacology & Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada*

ABSTRACT

Purpose The role of administrative databases for research on drug safety during pregnancy can be limited by their inaccurate assessment of the timing of exposure, as the gestational age at birth is typically unavailable. Therefore, we sought to develop and validate algorithms to estimate the gestational age at birth using information available in these databases.

Methods Using a population-based cohort of 286,432 mother–child pairs in British Columbia (1998–2007), we validated an ICD-9/10-based preterm-status indicator and developed algorithms to estimate the gestational age at birth on the basis of this indicator, maternal age, singleton/multiple status, and claims for routine prenatal care tests. We assessed the accuracy of the algorithm-based estimates relative to the gold standard of the clinical gestational age at birth recorded in the delivery discharge record.

Results The preterm-status indicator had specificity and sensitivity of 98% and 91%, respectively. Estimates from an algorithm that assigned 35 weeks of gestational age at birth to deliveries with the preterm-status indicator and 39 weeks to those without them were within 2 weeks of the clinical gestational age at birth in 75% of preterm and 99% of term deliveries.

Conclusions Subtracting 35 weeks (245 days) from the date of birth in deliveries with codes for preterm birth and 39 weeks (273 days) in those without them provided the optimal estimate of the beginning of pregnancy among the algorithms studied.

KEY WORDS—pharmacoepidemiology; pregnancy; premature birth; term birth; duration of pregnancy; claims databases; last menstrual period

Received 8 October 2011; Revised 13 January 2012; Accepted 19 March 2012

INTRODUCTION

Automated databases are commonly used in research on drug safety in pregnancy. These databases contain longitudinal data on health services utilization and pharmacy prescriptions or dispensing on a large number of individuals, thus permitting the study of rare exposures and outcomes.[1] Furthermore, they do not depend on retrospective recall, an important challenge for other data sources. On the other hand, they have a noteworthy limitation: the date of beginning of pregnancy is not routinely recorded.[2] Therefore, the gestational age at the time of maternal drug use is uncertain. Administrative databases contain, though, information that can support the estimation of the beginning of pregnancy.

In the absence of information on the beginning or the duration of pregnancy, several methods have been used to estimate them in automated databases. Often, researchers assumed a fixed duration of pregnancy of 270–280 days.[3,4] However, this is inaccurate for preterm and some term pregnancies. Some studies excluded suspected short gestations from the study population,[5,6] but this method is limited to the evaluation of outcomes unrelated to short gestation.[7] Other

*Correspondence to: A. V. Margulis, Harvard School of Public Health, Department of Epidemiology, 677 Huntington Avenue, Kresge Building, Boston, MA 02115, USA. E-mail: andreamargulis@post.harvard.edu

†Previous presentations: An abstract based on this research was presented as poster in the International Conference on Pharmacoepidemiology and Risk Management 2010 (abstract ID 709, A Claims-based Algorithm to Estimate the Date of the Last Menstrual Period). This research was part of the corresponding author's doctoral thesis; an earlier version of this manuscript has been printed to fulfill the University's requirements for graduation and was presented in the dissertation defense. The thesis is not available online.

methods estimated the beginning of pregnancy within a wide time window[8–10] (e.g., within 90 days before the first prenatal visit or other early pregnancy markers), introducing misclassification of the etiologically relevant timing of exposure.[7] In other studies, gestational age at birth was estimated from birth weight by using growth charts,[11,12] but this method requires the birth weight to be known and assumes infants follow the median growth trajectory. Because of the limitations of these methods, we set out to develop and validate algorithms to estimate the beginning of pregnancy by employing information that can be extracted from automated databases.

Using administrative data from a population-based cohort of pregnancies linked to gestational age information from birth discharge records, we first evaluated an indicator of preterm birth. Then, we created algorithms to estimate the gestational age at birth on the basis of the presence of the preterm birth indicator, maternal age, singleton/multiple status, and the timing of routine prenatal screening tests. Lastly, we compared our estimates, and the conventional method of assigning all pregnancies a duration of 280 days, with the gestational age at birth from clinical discharge records.

METHODS

Data source and study population

British Columbia provides health care through the British Columbia Medical Services Plan to over 94% of the population.[13] Patients' health care utilization is recorded in anonymized, linkable databases that include inpatient and outpatient diagnoses and procedures, health care provider visits, discharge records, dispensed prescriptions, and vital statistics. During the study period, diagnoses and procedures were coded in the International Classification of Diseases, ninth revision, Canada (ICD-9CA), ICD-10CA, and Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures. Other health services were coded in a fee-for-service coding system. Mothers and their offspring are linked by the British Columbia Ministry of Health in a perinatal database that has been used for research on reproductive health (96–97% of records are successfully linked).[14–17]

Unlike many administrative databases, our database contains a field with the gestational age at birth as clinically assessed and reported in the hospital discharge record. When there is no dating ultrasound, the gestational age at birth is assigned on the basis of the self-reported date of last menstrual period. When an early dating ultrasound is available and the due dates by self-reported date of last menstrual period and dating ultrasound are within 7 days of each other, the gestational age at birth is based on the self-reported date of last menstrual period. If the difference is larger, the gestational age at birth is based on the ultrasound. If neither of the two sources is considered reliable, the gestational ages at birth as reported by the caregiver in the birth form and by physical examination at birth are considered. This is the gestational age used in prenatal care and serves as the gold standard in our evaluation of alternative estimates of the gestational age at birth.

This study was approved by the Brigham and Women's Hospital institutional review board, and signed data use agreements were in place.

Our study population included all mother–child pairs with delivery date between October 1998 and March 2007. Because information was available for hospital births only, the sample was restricted to hospital deliveries (98% of all deliveries in British Columbia).[18] We required enrollment in the outpatient health care system for 365 + 280 days before delivery to ensure that the use of health care services during the year before gestation and the entire gestation would be recorded. Pregnancies with invalid gestational age at birth in the hospital discharge record (i.e., missing and shorter than 20 completed weeks or longer than 44 completed weeks) were excluded. Mother's age at delivery and the presence of maternal, pregnancy, and neonatal conditions that are associated with preterm deliveries were extracted from enrollment data and inpatient and outpatient claims from 645 days before to 60 days after delivery.

Preterm status

In this dataset, preterm deliveries can be identified on the basis of the standard clinical gestational age at birth, but researchers must usually resort to surrogate sources of information such as claims. We classified births as preterm in the presence of a claim for (1) ICD-9 codes 765 (*Disorders relating to short gestation and low birth weight*) or their ICD-10 approximately equivalent codes P05 (*Slow fetal growth and fetal malnutrition*) and P07 (*Disorders related to short gestation and low birth weight, not elsewhere classified*) and (2) ICD-9 644.0 and 644.2 (in 644, *Early or threatened labor*) or its ICD-10 approximately equivalent O60.1 (in O60, *Preterm labor*) in the first 60 days after delivery. We calculated the sensitivity, specificity, and positive and negative predictive values of these two definitions of preterm status, the combination of both, and their 95% confidence intervals. The reference was clinical standard gestational age at birth <37 completed weeks.

DOI: 10.1002/pds

Estimation of gestational age: conventional method and empirical modification

Conventional method: All pregnancies were assigned a fixed duration of 40 weeks, which is the median duration of human gestation[19] (hereafter, conventional method—40 weeks).

Modification to the conventional method: We assigned all pregnancies a fixed duration of 39 weeks, which is the median clinical gestational age at birth in the study population (hereafter, conventional method, empirical—39 weeks).

Estimation of gestational age at birth: new algorithms

We developed two groups of algorithms: algorithms based on our proposed claim-based preterm-status indicator and algorithms based on screening test claims. The first group, presented in the following section, builds on the conventional method but assigns different durations to preterm and term gestations.

Algorithms based on a claim-based preterm-status indicator. Algorithm A—preterms 36/terms 40 weeks: We assigned all preterm births a gestational age of 36 weeks, the most common gestational age at birth among preterm births,[20] and all non-preterm births a gestational age of 40 weeks.

Algorithm B—preterms 35/terms 40 weeks: We assigned all preterm births a gestational age of 35 weeks, which is the median gestational age at birth among the births with clinical gestational age at birth <37 completed weeks in the study population; and all non-preterm births a gestational age of 40 weeks, which is the median gestational age at birth among those with clinical gestational age at birth ≥37 completed weeks in the study population.

Algorithms based on screening test claims. These algorithms borrowed information from the pattern of claims for routine prenatal screening tests.[21] We first selected the pregnancy-specific screening tests indicated within narrow gestational age windows from British Columbia[22] and Canada[23–26] prenatal care guidelines (Table 1). We kept only the first occurrence of each code per pregnancy, under the assumption that first-time tests are indicated in a timely manner, whereas retests may legitimately occur after the intended gestational age window.

The gestational age at the moment of the test is unknown; therefore, we needed to estimate it (replicating the typical setting without information on gestational age at birth). To do this, we assigned each claim a gestational age equal to the midpoint of the test-specific gestational age window in British Columbia and Canada guidelines. We compared the midpoint and recommended gestational age window to the clinical gestational age distribution found in British Columbia perinatal database as the reference. The reference distribution was estimated as the number of days after the beginning of pregnancy, where the beginning of pregnancy was the date of delivery minus the clinical gestational age at birth. The guideline-based gestational age was accurate within days for a large proportion of subjects to the clinical gestational-age-at-birth-based gestational age (Table 1). In the algorithms in the succeeding paragraphs, the gestational age at the time

Table 1. Gestational age at the time of screening tests: guideline-recommended gestational age windows and data-driven pattern of use, in weeks^{days}, British Columbia, Canada, 1998–2007

Code description in British Columbia perinatal database	Screening test	Timing of screening tests: Guidelines Midpoint	Guidelines (range)	Clinical GAB-based Median	Clinical GAB-based (p25, p75)
Alpha fetoprotein	Serum integrated prenatal screen/integrated prenatal screen/quad screen[22]	16^2	(15^2, 17^0)	16^2	(15^4, 17^3)
Guided amniocentesis Amniocentesis and transabdominal Cytogenetic analysis—cultured amniotic fluid	Amniocentesis[23,25]	16^1	(15^0, 17^0)	16^1	(15^4, 17^3)
Obs.–B-scan (14 weeks or more)	Anatomical ultrasound[22,26]	19^1	(18^0, 20^0)	18^6	(17^5, 20^1)
Glucose, gestational assessment	Gestational diabetes screening[24]	26^1	(24^0, 28^0)	27^4	(26^1, 28^6)

GAB, gestational age at birth; p25, 25th percentile; p75, 75th percentile.

DOI: 10.1002/pds

of each screening test was assigned as the midpoint of the guideline recommendations.

Algorithm C—claim-based, average: We calculated one gestational age at birth per claim as [(date of delivery - date of claim) + midpoint of guideline-recommended gestational age window] for the screening tests in Table 1. For pregnancies with claims for two or more tests, we averaged the gestational ages at birth estimated from each of them.

Algorithm D—claim-based, regression: Clinical information can be incorporated to the estimation of gestational age at birth through the use of regression models. Because British Columbia perinatal database includes gestational age at birth, we were able to create linear regression models with gestational age at birth as the outcome, so that the estimated regression coefficients can be applied to other databases to estimate the beginning of pregnancy. The regression models included the following predictors: mother's age at delivery, our validated preterm-status indicator, multiple gestation, and the gestational age at birth assigned to each of the claims in Table 1. To create an algorithm that can be applied in most settings—where some claims may be missing for some pregnancies—and make use of all available data, we created a model for each possible combination of the four screening tests of interest; each pregnancy contributed only to the largest model it had complete data for. For example, the pregnancies with claims for all screening tests of interest were included in the largest model only. Thus, this algorithm comprised a set of 16 models that estimated a single gestational age at birth for each pregnancy. The coefficients and their standard errors were estimated on a randomly selected derivation set (50% of the study population); the predicted gestational ages at birth were calculated on the remaining 50% (validation set). Model specifications and a description of how to apply this are provided in the supporting information.

Algorithm E—claim-based, stratified regression: Because the timing of prenatal screening may differ between pregnancies that will end in a preterm delivery and those that will end in a term delivery, data were stratified on the basis of ICD 9/10-defined preterm status. The 16 linear regression models in algorithm D—claim-based, regression were run in each stratum (model specifications are provided in the supporting information).

Validation of estimated gestational age at birth against clinical gestational age at birth

For each algorithm, we calculated the proportion of pregnancies whose estimated gestational age at birth was within 1, 1+ to 2, 2+ to 4, or 4+ weeks of the clinical gestational age at birth recorded in the hospital discharge records, stratified by gestational age at birth <37 versus ≥ 37 completed weeks. We explored graphically the difference between the estimated and clinical gestational age at birth through histograms for selected methods.

All analyses were performed with SAS 9.1 (SAS Institute Inc., Cary, NC, USA) except the estimation of 95% confidence intervals for sensitivity, specificity, and positive and negative predictive values, which was performed with Episheet 2008.

RESULTS

We identified 209,532 women who gave birth to 286,968 newborns. We excluded 530 pregnancies with missing gestational age at birth and another six with gestational age at birth shorter than 20 completed weeks or longer than 44 completed weeks. The final study population comprised 286,432 newborns (84% of live births in British Columbia and 10% of live births in Canada).[27] The median gestational age at birth was 39 completed weeks (25th percentile, 38 weeks; 75th percentile, 40 weeks). There were 19,871 pregnancies (6.9%) that had gestational age at birth <37 completed weeks. The median gestational age at birth among them was 35 completed weeks (25th percentile, 34 weeks; 75th percentile, 36 weeks). On the basis of diagnostic codes, 24,396 pregnancies (8.5%) were classified as preterm. Participant characteristics are provided in Table 1 in the supporting information. There were 282,266 (98.5%) pregnancies that had at least one claim for a pregnancy-specific test listed in Table 1 during pregnancy.

The two preterm-status definitions based on ICD-9 code 765 *Disorders relating to short gestation and low birth weight* had a specificity of 98% and a sensitivity of 91%; all confidence intervals were narrow (Table 2). Their positive predictive value was 74%, and their negative predictive value was 99%. The definition that included claims for *Early or threatened labor* had a marginally higher sensitivity. Therefore, in subsequent analyses, we defined preterm status by the presence of claims for ICD-9 or ICD-10 codes for *Disorders relating to short gestation and low birth weight* or *Early or threatened labor*.

The gestational age at birth was estimable for all pregnancies by most methods. The exceptions were 4166 pregnancies without claims for the tests of interest, to which algorithm C—claim-based, average could not be applied, and one pregnancy, to which algorithm E—claim-based, stratified regression could not be applied (Table 3). The difference between the estimated

DOI: 10.1002/pds

Table 2. Validation of ICD-9/10-based definitions of preterm status: sensitivity, specificity, and positive and negative predicted values ($n = 286{,}432$), British Columbia, Canada, 1998–2007

	n (%)	Sensitivity	95%CI	Specificity	95%CI	PPV	95%CI	NPV	95%CI
Disorders relating to short gestation and low birth weight	24,245 (8.5)	0.91	0.90, 0.91	0.98	0.98, 0.98	0.74	0.74, 0.75	0.99	0.99, 0.99
Early or threatened labor	1,428 (0.5)	0.07	0.06, 0.07	1.00	1.00, 1.00	0.91	0.90, 0.93	0.94	0.93, 0.97
Disorders relating to short gestation and low birth weight *or* Early or threatened labor	24,396 (8.5)	0.91	0.91, 0.91	0.98	0.98, 0.98	0.74	0.74, 0.75	0.99	0.99, 0.99

CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

and the clinical gestational age at birth generally decreased and centered around 0 as more complex algorithms were applied (Figure 1). The methods that did not stratify on preterm status (i.e., conventional method—40 weeks and conventional method, empirical—39 weeks) overestimated the gestational age at birth of preterm gestations by over 3 weeks. Estimates by all methods were closer to the clinical gestational age at birth among term than among preterm gestations.

Among pregnancies with gestational age at birth <37 weeks, within-1-week agreement was highest for algorithm B—preterms 35/terms 40 weeks (68.1%, Table 3), whereas within-2-week agreement was highest for algorithm E—claim-based, stratified regression (75.8%). Among pregnancies with gestational age at birth ≥37 completed weeks, within-1-week and within-2-week agreements were highest for the conventional method, empirical—39 weeks (76.3% and 99.1%, respectively). Percents of agreement for other time intervals are provided in Table 3.

DISCUSSION

In this population-based analysis, algorithms based on claim-derived information provided good estimates of the gestational age at birth among preterm and term births and performed better than the conventional method.

The usefulness of these algorithms depends on the balance between the accuracy gained and the ease of implementation. Simply identifying the preterm births by using ICD-9/10 codes in claims and assigning them an appropriate gestational age at birth improved the accuracy of estimations dramatically in this subgroup. Assigning the term deliveries a gestational age of 39 weeks worked better than assigning them 40 weeks.

The preferred estimation method will depend upon the question at hand. The assignment of a gestational age of 35 weeks to preterm births and 39 weeks to term births maximized within-1-week agreement in our data, whereas using stratified regression for preterm births and assigning 39 weeks to term births maximized within-2-week agreement. Although researchers interested in short exposures (e.g., 1-week-long antibiotic therapies) may want to maximize within-1-week agreement of estimated and clinical gestational age at birth, in studies particularly sensitive to misclassification of preterm status, researchers may consider choosing methods to minimize the number of births for which the gestational age is overestimated (i.e., minimizing the sum of the three rightmost columns in Table 3).

The implementation of algorithm that consider a preterm status indicator is straightforward and requires only the identification of preterm deliveries from inpatient and outpatient claims and external information on the median gestational age at birth among preterm and term deliveries in the population of interest (use 35 and 39 if the latter is not available). A strength of this method is its robustness to late entry into prenatal care. Regression-based algorithms are somewhat more complex to implement; details on their implementation are provided in the supporting information.

The accuracy of the estimation of the gestational age at birth improved with the proposed algorithms among preterm deliveries, but it still remains lower as compared with term deliveries. Assigning a duration of pregnancy of 35 weeks will be inaccurate for the small percent of very short gestations. Evaluating exposures during specific pregnancy periods (e.g., antibiotics during the second gestational month) translates into differential misclassification of exposure if the outcome is associated with preterm status. The impact of misclassification would be larger for short-term exposures than from chronic ones. For example, we observed that first-trimester exposure to selective serotonin reuptake inhibitors (generally prescribed for chronic use) was more robust to the choice of beginning of pregnancy-estimating algorithm than exposure to fluconazole (episodic use; data not shown), as previously noted.[7]

This study was conducted on live births in administrative data by using information from records from mothers and offspring. Generalization to electronic medical record databases is straightforward, and at least one study employed a variant of one of our

DOI: 10.1002/pds

Table 3. Validation of estimated gestational age at birth against clinical gestational age, British Columbia, Canada, 1998–2007

			Estimated GAB is shorter than clinical GAB			Estimated and clinical GAB are within 1 week	Estimated GAB is longer than clinical GAB		
			4+ weeks	2+ to 4 weeks	1+ to 2 weeks		1+ to 2 weeks	2+ to 4 weeks	4+ weeks
	Algorithm	$n^{\dagger}$	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Pregnancies with clinical GAB <37 completed weeks									
	Conventional method—40 weeks	19,868						8,586 (43.2%)	11,282 (56.8%)
	Conventional method, empirical—39 weeks	19,868						12,736 (64.1%)	7,132 (3.9%)
A	Preterms 36/terms 40 weeks	19,868				11,052 (55.6%)	2,474 (12.5%)	3,829 (19.3%)	2,513 (12.7%)
B	Preterms 35/terms 40 weeks	19,868				13,256 (68.1%)	1,325 (6.7%)	3,093 (15.6%)	1,924 (9.7%)
C	Claim-based, average	19,365	1,240 (6.4%)	1,364 (7.0%)	1,817 (9.4%)	8,811 (45.4%)	2,961 (15.3%)	1,932 (10.0%)	1,240 (6.4%)
D	Claim-based, regression	9,953	9 (0.1%)	204 (2.1%)	840 (8.4%)	5,179 (52.0%)	1,438 (14.5%)	1,565 (15.7%)	718 (7.2%)
E	Claim-based, stratified regression	9,953	72 (0.1%)	334 (3.3%)	808 (8.1%)	5,172 (52.0%)	1,567 (15.7%)	1,497 (15.4%)	503 (5.1%)
Pregnancies with clinical GAB ≥37 completed weeks									
	Conventional method—40 weeks	266,564		119 (0.0%)	2,434 (1.0%)	197,589 (74.1%)	48,439 (18.2%)	17,983 (6.8%)	
	Conventional method, empirical—39 weeks	266,564	5 (0.0%)	2,548 (1.0%)	42,747 (16.0%)	203,281 (76.3%)	17,983 (6.8%)		
A	Preterms 36/terms 40 weeks	266,564	533 (0.2%)	2,439 (0.9%)	3,810 (1.4%)	196,823 (73.8%)	47,044 (17.8%)	15,915 (6.0%)	
B	Preterms 35/terms 40 weeks	266,564	1,669 (0.6%)	2,698 (1.0%)	4,483 (1.7%)	194,755 (73.1%)	47,044 (17.7%)	15,915 (6.0%)	
C	Claim-based, average	262,901	16,496 (6.3%)	23,332 (8.9%)	30,857 (11.7%)	131,163 (49.9%)	33,568 (12.8%)	18,007 (6.9%)	9,478 (3.6%)
D	Claim-based, regression	133,010	409 (0.3%)	3,714 (2.76%)	18,429 (13.9%)	88,723 (66.7%)	16,800 (12.6%)	4,497 (3.4%)	438 (0.3%)
E	Claim-based, stratified regression	133,009‡	272 (0.2%)	2,846 (2.1%)	19,258 (14.5%)	87,925 (66.1%)	17,667 (13.3%)	4,642 (3.5%)	399 (0.3%)

GAB, gestational age at birth.

†Regression results from validation set.

‡No preterm pregnancies in the derivation set had amniocentesis-related claims only; therefore, the GAB could not be predicted for the one such pregnancy in the validation set and this pregnancy could not be incorporated in the analysis.

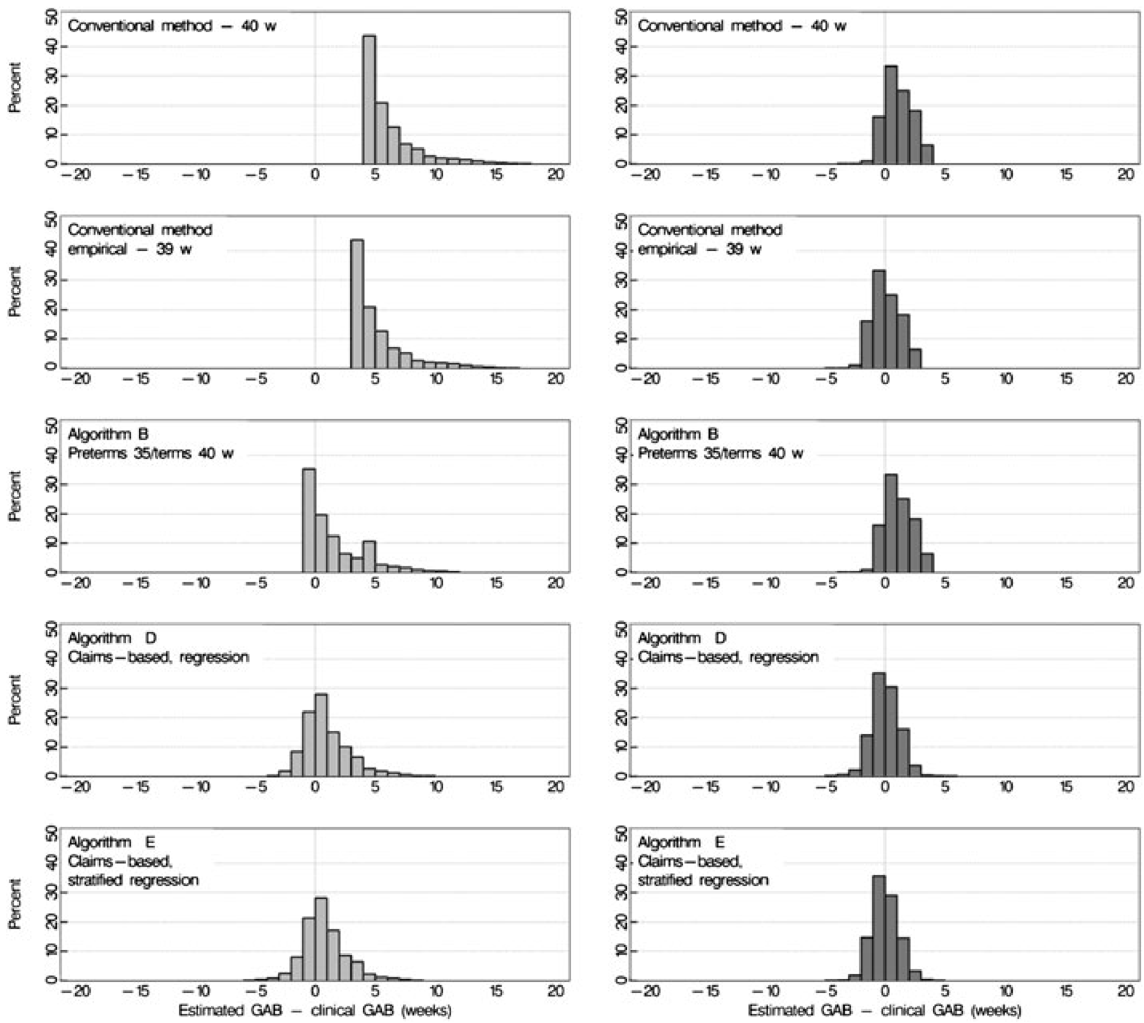

Figure 1. Distribution of estimated gestational age at birth minus clinical gestational age at birth among pregnancies with gestational age at birth <37 completed weeks (left column) and ≥37 completed weeks (right column), in weeks. Negative numbers on the horizontal axis represent estimated gestational age at birth shorter than clinical gestational age at birth; positive numbers represent estimated gestational age at birth longer than clinical gestational age at birth. GAB, gestational age at birth

proposed algorithms,[28] but further research is needed before recommendations can be made for the estimation of the length of gestation in stillbirths and abortions. Our methods are based on the use of ICD-9/10 codes for short gestation and early labor to identify preterm deliveries and are thus generalizable to databases with comparable coding practices. The positive predictive value of the preterm indicator might increase by using more restrictive ICD 9/10 codes, at the risk, though, of a potentially decreased sensitivity. Furthermore, in other databases, information on the gestational age at birth may be retrieved from the fifth digit in the ICD 9 code 765.2x. However, the fifth digit is not recorded in British Columbia perinatal database. If the fifth digit is available, methods that make use of such information should be considered, possibly after a validation study. Regression results may be optimistic because the derivation and validation datasets come from the same source. Two regression models in algorithm D and six in algorithm E are based on a small number of pregnancies (less than 10 pregnancies per predictor); a few regression coefficients and/or their standard errors in those strata are not estimable because of sparse data. Also, we assumed all pregnancies were independent observations, although included in the study population are multifetal gestations and siblings. As a result, standard errors of the regression coefficients may be inaccurate. Regression results are applicable to other populations with characteristics that may affect the duration of pregnancy (e.g., ethnic background) and patterns of prenatal care similar to the ones in British Columbia. Prenatal care recommendations in the

DOI: 10.1002/pds

USA[29–31] and the UK[32,33] are comparable with the ones in British Columbia. Otherwise, coefficients would need to be adapted to the local guidelines. It should be noted that the clinical gestational age at birth in this database is recorded in completed weeks; thus, 1 week is the maximum precision attainable in this study. Refinement of the preterm indicator could involve incorporating codes for the postnatal care of preterm infants; a post-term indicator may also be considered.

In conclusion, subtracting 35 weeks (245 days) from the birth date in deliveries with preterm-related codes and 39 weeks (273 days) in deliveries without them provided optimal estimates of the beginning of pregnancy in terms of ease of implementation and accuracy. This method can be implemented in mother–offspring linked data for drug safety in pregnancy and related research.

CONFLICT OF INTEREST

The authors have declared that there is no conflict of interest.

KEY POINTS

- A variety of methods are currently used to estimate the beginning of pregnancy in automated databases.
- We validated simple algorithms that improve on commonly used methods.
- These algorithms make use of information available in automated databases.
- Estimates are more accurate among term deliveries than among preterm deliveries.

ACKNOWLEDGEMENTS

This work was supported by NICHD grant R21HD055479 (PI: Soko Setoguchi).

The Pharmacoepidemiology Program at the Harvard School of Public Health is supported by training grants from Pfizer, Novartis, and Asisa.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix Section 1: Table 1. Participant Characteristics, British Columbia, Canada, 1998-2007.

Appendix Section 2: How to implement claim-based regression algorithms.

Appendix Section 2: Table 2. Parameter Estimates for Claims-based Regression Algorithms in the Derivation Set, Standard Errors, R^2 and adjusted R^2, British Columbia, Canada, 1998-2007.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

1. Andrews EB, Tennis P. Promise and pitfalls of administrative data in evaluating pregnancy outcomes. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1181–1183. DOI: 10.1002/pds.1499
2. Manson JM, McFarland B, Weiss S. Use of an automated database to evaluate markers for early detection of pregnancy. *Am J Epidemiol* 2001; **154**: 180–187.
3. Raebel MA, Ellis JL, Andrade SE. Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data. *Pharmacoepidemiol Drug Saf* 2005; **14**: 829–836.
4. Johnsen SL, Wilsgaard T, Rasmussen S, *et al*. Fetal size in the second trimester is associated with the duration of pregnancy, small fetuses having longer pregnancies. *BMC Pregnancy Childbirth* 2008; **8**: 25. DOI: 1471-2393-8-25 [pii] 10.1186/1471-2393-8-25.
5. Andrade SE, Raebel MA, Morse AN, *et al*. Use of prescription medications with a potential for fetal harm among pregnant women. *Pharmacoepidemiol Drug Saf* 2006; **15**: 546–554.
6. Andrade SE, McPhillips H, Loren D, *et al*. Antidepressant medication use and risk of persistent pulmonary hypertension of the newborn. *Pharmacoepidemiol Drug Saf* 2009; **18**: 246–252.
7. Toh S, Mitchell AA, Werler MM, *et al*. Sensitivity and specificity of computerized algorithms to classify gestational periods in the absence of information on date of conception. *Am J Epidemiol* 2008; **167**: 633–640.
8. Hardy JR, Leaderer BP, Holford TR, *et al*. Safety of medications prescribed before and during early pregnancy in a cohort of 81,975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2006; **15**: 555–564.
9. Cole JA, Ephross SA, Cosmatos IS, *et al*. Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1075–1085. DOI: 10.1002/pds.1463
10. Cole JA, Modell JG, Haight BR, *et al*. Bupropion in pregnancy and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007; **16**: 474–484. DOI: 10.1002/pds.1296
11. Cooper WO, Hernandez-Diaz S, Arbogast PG, *et al*. Major congenital malformations after first-trimester exposure to ACE inhibitors. *N Engl J Med* 2006; **354**: 2443–2451.
12. Cooper WO, Willy ME, Pont SJ, *et al*. Increasing use of antidepressants in pregnancy. *Am J Obstet Gynecol* 2007; **196**: 544 e1-544 e5. DOI: S0002-9378(07) 00144-5 [pii] 10.1016/j.ajog.2007.01.033
13. Hanley GE, Morgan S. On the validity of area-based income measures to proxy household income. *BMC Health Serv Res* 2008; **8**: 79. DOI: 1472-6963-8-79 [pii] 10.1186/1472-6963-8-79.
14. Oberlander TF, Warburton W, Misri S, *et al*. Neonatal outcomes after prenatal exposure to selective serotonin reuptake inhibitor antidepressants and maternal depression using population-based linked health data. *Arch Gen Psychiatry* 2006; **63**: 898–906.
15. Oberlander TF, Warburton W, Misri S, *et al*. Effects of timing and duration of gestational exposure to serotonin reuptake inhibitor antidepressants: population-based study. *Br J Psychiatry* 2008; **192**: 338–343.
16. Oberlander TF, Warburton W, Misri S, *et al*. Major congenital malformations following prenatal exposure to serotonin reuptake inhibitors and benzodiazepines using population-based health data. *Birth Defects Res B Dev Reprod Toxicol* 2008; **83**: 68–76.
17. Hanley GE, Janssen PA, Greyson D. Regional variation in the cesarean delivery and assisted vaginal delivery rates. *Obstet Gynecol* 2010; **115**: 1201–1208. DOI: 10.1097/AOG.0b013e3181dd918c 00006250-201006000-00016 [pii]
18. Statistics Canada. Table 102–4516 - Live births and fetal deaths (stillbirths), by place of birth (hospital and non-hospital), Canada, provinces and territories, annual. http://cansim2.statcan.gc.ca/cgi-win/cnsmcgi.exe?Lang=E&CNSM-Fi=CII/CII_1-eng.htm (accessed June 15, 2010).
19. Bergsjo P, Denman DW, 3rd, Hoffman HJ, *et al*. Duration of human singleton pregnancy. A population-based study. *Acta Obstet Gynecol Scand* 1990; **69**: 197–207.
20. Reddy UM, Ko CW, Raju TN, *et al*. Delivery indications at late-preterm gestations and infant mortality rates in the United States. *Pediatrics* 2009; **124**: 234–240. DOI: 124/1/234 [pii] 10.1542/peds.2008-3232.

21. Walker AM. Pattern recognition in health insurance claims databases. *Pharmacoepidemiol Drug Saf* 2001; **10**: 393–397. DOI: 10.1002/pds.611 [doi]
22. Guideline: prenatal screening for Down syndrome, trisomy 18 and open neural tube defects. BC Prenatal Genetic Screening Program. www.bcprenatalscreening.ca (accessed August 15, 2009).
23. Chodirker BN, Cadrin C, Davies GA, *et al*. Canadian Guidelines for Prenatal Diagnosis: techniques of prenatal diagnosis. *J Obstet Gynaecol Can* 2001; **105**: 1–9.
24. Berger H, Crane J, Farine D, *et al*. Screening for gestational diabetes mellitus. *J Obstet Gynaecol Can* 2002; **24**: 894–912.
25. Wilson RD, Davies G, Gagnon A, *et al*. Amended Canadian guideline for prenatal diagnosis (2005) change to 2005-techniques for prenatal diagnosis. *J Obstet Gynaecol Can* 2005; **27**: 1048–1062.
26. Summers AM, Langlois S, Wyatt P, *et al*. Prenatal screening for fetal aneuploidy. *J Obstet Gynaecol Can* 2007; **29**: 146–179.
27. Statistics Canada. Table 102–4512 - Live births, by weeks of gestation and sex, Canada, provinces and territories, annual, CANSIM (database). http://cansim2.statcan.gc.ca/cgi-win/CNSMCGI.PGM?Lang=Eng&Dir-Rep=&CNSM-Fi=CII_1-eng.htm (accessed March 25, 2009).
28. Alonso A, Jick SS, Olek MJ, *et al*. Recent use of oral contraceptives and the risk of multiple sclerosis. *Arch Neurol* 2005; **62**: 1362–1365. DOI: 62/9/1362 [pii] 10.1001/archneur.62.9.1362
29. Chapter 4 - Antepartum care in: Guidelines for Perinatal Care (AAP/ACOG), 6th edition. 2007. http://www.acog.org/~/media/Guidelines%20for%20Perinatal%20Care/Antepartum%20Care.ashx?dmc=1&ts=20111227T0946493494 (accessed December 25, 2011).
30. ACOG Practice Bulletin. Clinical management guidelines for obstetrician-gynecologists. Number 30, September 2001 (replaces Technical Bulletin Number 200, December 1994). Gestational diabetes. *Obstet Gynecol* 2001; **98**: 525–538.
31. Goldberg JD. Routine screening for fetal anomalies: expectations. *Obstet Gynecol Clin North Am* 2004; **31**: 35–50. DOI: 10.1016/s0889-8545(03)00118-9
32. Antenatal care: routine care for the healthy pregnant woman. Clinical Guideline. 2008. RCOG Press. http://www.nice.org.uk/nicemedia/live/11947/40145/40145.pdf (accessed December 25, 2011).
33. Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Quick reference guideline. 2008. http://www.nice.org.uk/nicemedia/live/11946/41321/41321.pdf (accessed December 25, 2011).