Practice of Epidemiology

Sensitivity and Specificity of Computerized Algorithms to Classify Gestational Periods in the Absence of Information on Date of Conception

Sengwee Toh¹, Allen A. Mitchell², Martha M. Werler², and Sonia Hernández-Díaz¹

¹ Department of Epidemiology, Harvard School of Public Health, Boston, MA.
² Slone Epidemiology Center at Boston University, Boston, MA.

Received for publication June 20, 2007; accepted for publication November 8, 2007.

To evaluate the accuracy of computerized algorithms for pinpointing periods of exposure to medications during pregnancy in the absence of data on timing of conception, the authors used data from a population-based sample of nonmalformed infants in the Slone Epidemiology Center Birth Defects Study in 1998–2006 (United States and Canada; N = 3,177). The standard was defined as any antiinfective use from 2 weeks after the last menstrual period through the third gestational month, which was compared with results obtained after defining the beginning of pregnancy as either 270 days before the birth date (delivery-date algorithm) or the date of the first prenatal visit (pregnancy-indicator algorithm). The sensitivity was 92% (95% confidence interval: 88, 95) for the delivery-date algorithm and 59% (95% confidence interval: 53, 65) for the pregnancy-indicator algorithm. The specificity was higher than 98% for both algorithms. The sensitivity for the delivery-date algorithm among women with preterm births was 66% (95% confidence interval: 49, 80). For women without pregnancy complications, subtraction of 270 days from the delivery date might be accurate for timing first-trimester prescription drug use in automated databases. However, the sensitivity of this algorithm is lower for preterm deliveries, suggesting limited validity to assess drug safety for pregnancy outcomes associated with prematurity.

Abbreviations: BDS, Slone Epidemiology Center Birth Defects Study; LMP, last menstrual period.

Editor’s note: An invited commentary on this article appears on page 641, and the authors’ response is published on page 644.

The safety of medication use during pregnancy has profound public health implications. Although presumably only a small proportion of fetal malformations is induced by medications, the absolute number of excess cases can be high given that 120,000 children are born with birth defects each year in the United States alone (1). While the teratogenicity of certain drugs (e.g., isotretinoin) has been well identified, the safety profile for most others remains largely unknown, and less teratogenic but more widely used drugs may be associated with an even higher number of infants with birth defects (2).

Premarking clinical trials normally exclude pregnant women for ethical reasons (2, 3). As a result, when a new drug becomes available to the general population, information regarding its safety during pregnancy is limited and is based largely on animal studies (2, 3). Unfortunately, human teratogenesis usually cannot be directly extrapolated from animal studies. Postmarketing observational studies thus become a crucial way to evaluate the safety of in utero exposure to medications in humans. However, these studies are often limited by small numbers of exposed women and/or outcomes under study.

Correspondence to Sengwee Toh, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: swtoh@hsph.harvard.edu).
There is an increasing trend in the use of automated databases to study the utilization and safety of drugs during pregnancy; examples include the computerized administrative databases in the HMO Research Network Center for Education and Research on Therapeutics (4) and the automated medical records in the General Practice Research Database (5). Several attractive features of automated databases render them a potentially powerful data source in this research area because they often include information on a large number of pregnant women, with detailed prescription records for mothers and clinical information on both mothers and infants. In addition, data are collected prospectively and are thus free from recall bias.

However, these databases are not without their limitations. The study of drug safety during pregnancy requires that data sets have mother-child linkage capability. Moreover, since the first trimester, during which organogenesis takes place, is the most etiologically relevant period with regard to teratogenesis, accurate estimation of the onset of gestation is essential in birth defects research. Automated databases often lack the necessary information, such as last menstrual period (LMP) and gestation length, to establish timing of conception. In the absence of these data, several computer algorithms have been developed to identify trimesters of pregnancy. Although mostly study specific, these algorithms share certain characteristics, such as using delivery date as the reference date to classify gestational periods. Unfortunately, these algorithms are rarely validated.

Taking advantage of a data source with detailed information on medication use and date of the LMP, we assessed the accuracy of computerized algorithms for determining whether prescription medications were used during the first trimester. Our assessment was based on timing of exposure to antifungal agents, the most common class of prescription medications in pregnancy.

**MATERIALS AND METHODS**

**Data source**

We used data from the Slone Epidemiology Center Birth Defects Study (BDS), a multicenter case-control surveillance study of birth defects in relation to environmental exposures (particularly medications). More than 35,000 mothers of babies with and without birth defects from the greater metropolitan areas of Boston, Massachusetts; Philadelphia, Pennsylvania; Toronto, Canada; and San Diego, California, as well as selected regions in Iowa, Massachusetts, and New York State, have been interviewed since 1976. Study subjects are identified through review of admissions and discharges at major pediatric referral hospitals and clinics, at birth hospitals, and in logbooks in neonatal intensive care units; through weekly telephone contact with collaborators at newborn nurseries in community hospitals; and through collaborations with state birth defects registries. Since 1998, the study has also included a random sample of Massachusetts births. Institutional review board approval was obtained from each of the participating institutions, and mothers provided informed consent before participation. The study is fully compliant with requirements of the Health Insurance Portability and Accountability Act (HIPAA), and requests for waivers of authorization for access to protected health information are submitted as appropriate. The current study was restricted to the population-based sample of 3,177 nonmalformed newborns from Massachusetts whose mothers were ascertained by BDS between 1998 and 2006.

**Timing of pregnancy periods**

The BDS collects information on LMP, whether based on maternal recall or ultrasound examination. The LMP date allows estimation of the due date, approximate conception date, and gestational age at birth. In this study, we defined the first day of conception as 14 days after the LMP and the first trimester of pregnancy as 90 days following the day of conception.

**Exposure ascertainment and definition**

As part of the BDS, telephone interviews are conducted by trained nurses within 6 months of delivery. Both mothers and nurses are unaware of study hypotheses at the time of interview. Using structured questionnaires, the interviewers systematically collect data on demographic, reproductive, lifestyle, and medical factors; diet; and medication use from 2 months before conception through the entire pregnancy. To facilitate recall, medication use is identified by asking a series of increasingly detailed questions, including standardized questions on drugs used for a list of indications (e.g., infections), specific conditions (e.g., sinus infection), and specific drug products (e.g., amoxicillin). This approach has been shown to identify an additional 25–40 percent of users who would have been misclassified as nonusers when a less detailed questionnaire is used (6). When possible, reported medications are verified by having the mother read information from the medication package. To identify timing of drug exposure, a 12-month calendar covering periods before and after pregnancy is used. Special dates (e.g., LMP, major holidays) are marked to help enhance recall. Medication data are carefully collected on starting and stopping dates, duration, frequency, indication, form, and number of pills per day.

For the current study, we considered a woman exposed if she reported having ever used an antifungal during the first 3 months after the estimated date of conception. We performed alternative analyses by classifying exposure status as at least 3 days and at least 7 days of use.

**Algorithms commonly used to classify trimesters during pregnancy when information on conception time has not been systematically recorded**

We evaluated two algorithms used in recently published studies.

Delivery-date algorithm. Andrade et al. (4) used the delivery date as the reference date and defined the first day of the exposure period (i.e., proxy for the first day of gestation) as the delivery date minus 270 days; each trimester was then assigned a 90-day interval. To avoid misclassification of
exposure in shorter gestations, they excluded women with documented conditions associated with preterm birth, including preeclampsia, hypertension, diabetes, early onset of labor, and intrauterine death.

Pregnancy-indicator algorithm. Hardy et al. (7) used an algorithm in which the first indication of pregnancy (i.e., a prenatal visit, positive pregnancy test, or any pregnancy-related diagnosis or procedure) recorded in the database was the reference date and defined this date as the index day for exposure. They considered two periods of exposure: the 90 days prior to the reference date, which presumably captured the actual date of conception, and the 70 days after that date, as a proxy for the first trimester. To reduce misclassification of early pregnancy exposure among women with late prenatal care or preterm deliveries, they excluded women whose prenatal visits occurred over a period of less than 7 months, as well as those whose first recorded pregnancy event and ending records (e.g., births, spontaneous abortion) spanned 280 days or more (7).

Main analysis

In the BDS data, we replicated as closely as possible the algorithms used by Andrade et al. (delivery-date algorithm) (4) and Hardy et al. (pregnancy-indicator algorithm) (7). In addition, to provide the “best case” scenario, we restricted our analyses to women giving birth to term liveborns with no abnormality (N = 2,949). A term delivery was defined as a birth at a gestational age of at least 37 weeks after the LMP. To assess the performance of these algorithms in the presence of preterm deliveries, we also conducted our analyses by considering all women (N = 3,177).

To apply the delivery-date algorithm, we defined the first day of the exposure period as the delivery date minus 270 days and considered the first 90 days to define exposure. Following this algorithm, we excluded from the BDS population of all nonmalformed liveborns those pregnancies affected by conditions associated with preterm birth (n = 1,228): preeclampsia, hypertension, diabetes, multiple gestation, early onset of labor or use of tocolytics, and infertility treatments.

To apply the pregnancy-indicator algorithm, we defined the first day of the exposure period as the date of the first prenatal visit and considered the 70 days after that date to define exposure. The date of the first prenatal visit had been collected for only a period of time in the BDS, whereas the date of the first positive pregnancy test had been consistently collected. On the basis of those women who provided information on both dates during the interview, we calculated the average time from first pregnancy test to first prenatal visit (18 days); for those women without a date for the first prenatal visit, we estimated that date by adding 18 days to the date of the pregnancy test. Following this algorithm, we excluded pregnancies with fewer than 7 months between the mother’s first prenatal visit and the infant’s birth date, and those with more than 280 days between the first prenatal visit and delivery date (n = 730).

We considered as the “gold standard” the classification of first-trimester exposure to antifungal drugs based on the full data on timing of pregnancy systematically collected by the BDS. We compared these results with those obtained through the automated algorithms and calculated 1) the proportion of women exposed during the first trimester according to the standard who were also classified as exposed by the algorithm (i.e., sensitivity); and 2) the proportion of women unexposed during the first trimester according to the standard who were also classified as unexposed by the algorithm (i.e., specificity). Exact 95 percent confidence intervals for the sensitivity and specificity were also estimated. All analyses were performed with SAS 9.1 software (SAS Institute, Inc., Cary, North Carolina).

Impact of exposure misclassification on relative risk estimates

To quantify the implications of inaccurate timing of pregnancy on misclassification of in utero antifungal use during the first trimester, we used the method proposed by Greenland (8) to estimate the impact of classification errors of dichotomous variables. Using the sensitivities and specificities estimated in our analyses, as well as the prevalence of antifungal use obtained from maternal interviews in the BDS, we estimated the magnitude of information bias by comparing hypothetical relative risks calculated from complete and incomplete data on timing of conception. Despite the prospective nature of data collection in automated databases, algorithms based on factors subsequent to the event of interest might introduce differential misclassification of prenatal exposures with respect to the outcome if such factors are associated with the outcome of interest. For example, if the accuracy of the algorithm depends on gestational age at birth, misclassification of first-trimester antifungal use may be different for infants with birth defects, who are more frequently preterm, than for healthy newborns. On the other hand, if the algorithm used to define timing of exposure is based on factors not associated with the event of interest (e.g., based on first pregnancy visit and the date of such visit is not associated with the specific pregnancy outcome), then the misclassification would be nondifferential. We therefore considered both differential and nondifferential misclassification of exposure.

RESULTS

The prevalence of antifungal use during the first trimester based on maternal recall was 12.5 percent for the 3,177 mothers of nonmalformed liveborns and 12.2 percent for the 2,949 with full-term deliveries. Most antifungals were used for 7–14 days (53 percent), with penicillins and antifungals being the most commonly reported drugs.

Characteristics of the population

Selected characteristics of study subjects are shown in table 1. For the delivery-date algorithm, excluding women with pregnancy complications reduced the sample size to 1,949 (14 percent of the excluded and 3 percent of the included women had preterm deliveries). For the pregnancy-indicator algorithm, restricting the population to women with at least 7 months and less than 280 days between the
first visit and the birth date reduced the sample to 2,447 women (27 percent of the excluded and 1 percent of the included women had preterm deliveries). Overall, the first day of the exposure period estimated by the delivery-date algorithm was closer to the date of conception than that estimated by the pregnancy-indicator algorithm (table 1).

Sensitivity and specificity of each algorithm

The sensitivity of the delivery-date algorithm exceeded that of the pregnancy-indicator algorithm across different definitions of exposure (table 2). Both algorithms had high specificities, which increased when the definitions of "exposed" required longer durations of antiinfective use. If the pregnancy-indicator algorithm had used a 90-day period (i.e., adding 20 days to the original 70-day interval), the sensitivity and specificity for any use would have been 60.3 (95 percent confidence interval: 54.7, 65.8) and 96.1 (95 percent confidence interval: 95.1, 96.8) percent, respectively. When we used the date of the first pregnancy test as the first pregnancy indicator (with a 70-day definition), the sensitivity and specificity of the pregnancy-indicator algorithm for any use was 74.9 (95 percent confidence interval: 70.1, 79.3) and 99.4 (95 percent confidence interval: 98.9, 99.6) percent, respectively.

Impact of preterm births on algorithm performance

When we restricted the sample to women with term deliveries rather than applying the exclusion criteria imposed by the algorithms, the sensitivity was higher than when mothers of preterm liveborns were included (table 3). In the extreme, when we restricted our analyses to women with preterm deliveries (n = 228), the sensitivity for each algorithm was much lower than that estimated for women with term deliveries only (table 3). The specificity was higher than 97 percent for both algorithms, and the confidence intervals overlapped.

Impact on relative risk estimates

In our sensitivity analyses, we considered a population in which 10 percent of the infants were exposed during the first trimester (e.g., to antibiotics), and we hypothesized that

| TABLE 1. Selected characteristics of women recruited in Massachusetts who gave birth to liveborns with no abnormality, the Stone Epidemiology Center Birth Defects Study, 1998–2006 |
|---|---|---|---|---|
| Characteristic | No.* | Mean (SD†) | Median | Minimum | Maximum |
| All women (N = 3,177) | | | | | |
| Maternal age (years) | 3,177 | 30.1 (5.9) | 31.0 | 13.0 | 46.0 |
| Gestation length (days from conception) | 3,177 | 261.6 (12.0) | 263.0 | 150.0 | 321.0 |
| Conception date minus the first day of the exposure period‡ | | | | | |
| Delivery-date algorithm (days) | 3,177 | −8.4 (12.0) | −7.0 | −120.0 | 51.0 |
| Pregnancy-indicator algorithm (days) | 3,097 | 40.8 (18.0) | 37.0 | 1.0 | 224.0 |
| Women with term births only (n = 2,949)§ | | | | | |
| Gestation length (days from conception) | 2,949 | 263.9 (8.4) | 264.0 | 245.0 | 321.0 |
| Conception date minus the first day of the exposure period‡ | | | | | |
| Delivery-date algorithm (days) | 2,949 | −6.1 (8.4) | −6.0 | −25.0 | 51.0 |
| Pregnancy-indicator algorithm (days) | 2,877 | 40.7 (17.5) | 37.0 | 3.0 | 215.0 |
| Women included in the delivery-date algorithm analysis (n = 1,949)¶ | | | | | |
| Gestation length (days from conception) | 1,949 | 263.9 (9.4) | 265.0 | 150.0 | 296.0 |
| Conception date minus the first day of the exposure period by this algorithm (days)‡ | 1,949 | −6.1 (9.4) | −5.0 | −120.0 | 26.0 |
| Women included in the pregnancy-indicator algorithm analysis (n = 2,447)¶ | | | | | |
| Gestation length (days from conception) | 2,447 | 264.4 (8.5) | 265.0 | 217.0 | 321.0 |
| Conception minus the first day of the exposure period in this algorithm (days)‡ | 2,441 | 35.5 (9.7) | 35.0 | 1.0 | 86.0 |

* The numbers of women within each category might be different because of missing information during the data collection process (e.g., patients did not answer or refused to answer a particular question).
† SD, standard deviation.
‡ The estimated conception date was defined as 14 days after the last menstrual period. In the delivery-date algorithm, the first day of the exposure period was 270 days before the date of delivery. In the pregnancy-indicator algorithm, the first prenatal visit was used as the first day of the exposure period.
§ Term birth was defined as birth after a gestational length of at least 37 weeks from the last menstrual period.
¶ The delivery-date algorithm restricted the population to pregnancies not affected by conditions associated with preterm delivery (1,228 women excluded). The pregnancy-indicator algorithm restricted the population to pregnancies with at least 7 months and less than 280 days between the first pregnancy visit and delivery (730 women excluded).
1 percent of the unexposed and 2 percent of the exposed had the outcome of interest (e.g., cardiovascular malformation), for an estimated relative risk of 2.0. In this paper, results are also presented for hypothetical relative risks of 1.5 and 3.0. We calculated the relative risk estimate that would have been observed under the assumption that the sensitivity ranged from 75 percent to 100 percent and the specificity was 99 percent for all women (figure 1). Realistic estimates based on our results assumed perfect ascertainment of exposure but imperfect timing. In figure 2, we also considered the scenario in which the proportion of false negatives was two times higher among newborns with the event (e.g., because the event was associated with preterm delivery).

Findings show that, with nondifferential misclassification, the biased relative risk estimates would be slightly closer to the null than the true ones. For example, with a sensitivity of 80 percent, we would have observed a relative risk of 1.86 rather than a true relative risk of 2.0. It was also apparent that, for relatively infrequent exposures, the impact of false positives (i.e., low specificity) was greater than the impact of false negatives (i.e., low sensitivity). However, with differential misclassification, the relative risk estimates would be substantially biased even under plausible scenarios of exposure misclassification. For example, with a sensitivity of 70 percent for cases and 85 percent for noncases, we would have observed a relative risk of 1.57 rather than a true relative risk of 2.0.

### TABLE 2. Classifications of antiinfective use during the first trimester by each algorithm against the “gold standard” under various exposure definitions for women in the Slone Epidemiology Center Birth Defects Study, 1998–2006*

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Gold standard</th>
<th>Any use (≥1 day)</th>
<th>≥3 days</th>
<th>≥7 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Use</td>
<td>Nonuse</td>
<td>Use</td>
</tr>
<tr>
<td>Delivery-date algorithm† (n = 1,949)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td>198</td>
<td>8</td>
<td>180</td>
</tr>
<tr>
<td>Nonuse</td>
<td></td>
<td>17</td>
<td>1,726</td>
<td>19</td>
</tr>
<tr>
<td>Sensitivity (%) and 95% CI‡</td>
<td></td>
<td>92.1</td>
<td>87.6, 95.3</td>
<td>90.5</td>
</tr>
<tr>
<td>Specificity (%) and 95% CI‡</td>
<td></td>
<td>99.5</td>
<td>99.1, 99.8</td>
<td>99.7</td>
</tr>
<tr>
<td>Pregnancy-indicator algorithm§ (n = 2,447)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td>186</td>
<td>40</td>
<td>165</td>
</tr>
<tr>
<td>Nonuse</td>
<td></td>
<td>129</td>
<td>2,086</td>
<td>125</td>
</tr>
<tr>
<td>Sensitivity (%) and 95% CI‡</td>
<td></td>
<td>59.1</td>
<td>53.3, 64.5</td>
<td>56.9</td>
</tr>
<tr>
<td>Specificity (%) and 95% CI‡</td>
<td></td>
<td>98.1</td>
<td>97.5, 98.7</td>
<td>98.3</td>
</tr>
</tbody>
</table>

* Women who used an antiinfective for at least x days (x = 1, 3, or 7) within the period of interest were classified as exposed, and as nonexposed otherwise. The numbers may not add to totals because of missing data.
† The beginning of the exposure period was defined as 270 days before the date of delivery.
‡ Exact 95% confidence interval (CI).
§ The beginning of the exposure period was defined as the date of the first prenatal visit.

### TABLE 3. Sensitivity and specificity of two algorithms developed to classify drug use during the first trimester of pregnancy by gestation length, the Slone Epidemiology Center Birth Defects Study, 1998–2006*

<table>
<thead>
<tr>
<th>Any antiinfective use</th>
<th>Delivery-date algorithm†</th>
<th>Pregnancy-indicator algorithm§</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI‡</td>
</tr>
<tr>
<td>All women (N = 3,177)</td>
<td>90.0</td>
<td>86.6, 92.7</td>
</tr>
<tr>
<td>Women with term births only (n = 2,949)†</td>
<td>92.5</td>
<td>89.3, 95.0</td>
</tr>
<tr>
<td>Women with preterm births only (n = 228)¶</td>
<td>65.8</td>
<td>48.7, 80.4</td>
</tr>
</tbody>
</table>

* Women who used an antiinfective for at least 1 day were classified as exposed, and as nonexposed otherwise.
† The beginning of the exposure period was defined as 270 days before the date of delivery.
‡ Exact 95% confidence interval (CI).
¶ Term birth was defined as birth after a gestational length of at least 37 weeks from the last menstrual period.
DISCUSSION

In the absence of information on timing of conception, pinpointing first-trimester exposures using the date of delivery as the reference date and subtracting 270 days resulted in greater sensitivity and slightly better specificity than using the first pregnancy-indicating event identified in automated databases as the reference date. Not surprisingly, the accuracy of the former algorithm was better for women who delivered at term than for those with preterm deliveries.

Delivery-date algorithm

Women conceiving normally (e.g., no maternal complication, no fetal abnormality) usually have a fairly stable length of gestation. Consequently, upon exclusion of pregnancy complications and preterm deliveries, the estimated day of conception extrapolated from the date of delivery is fairly consistent and close to the actual date of conception. Therefore, algorithms based on date of birth can provide acceptable estimates of drug utilization during specific pregnancy periods. However, such exclusions come with both challenges and potential biases. First, in the absence of information on gestational age, surrogates for preterm delivery have to be used in databases, which results in exclusion of about 40 percent of all pregnancies, many of them term or close to term. Such restrictions could reduce the representativeness of the study population with respect to the frequency and patterns of drug prescription during pregnancy. For instance, results would underestimate the prescribing of drugs used to treat the conditions excluded from the analyses as well as other conditions associated with them. Second, if the objective is to study the risk of pregnancy complications (e.g., birth defects), which are often associated with gestational age at birth, such events have to be included. Yet, if pregnancy complications and preterm deliveries are included, the differential misclassification of exposure for term and preterm infants might result in substantially biased relative risk estimates, as we discuss below.

Pregnancy-indicator algorithm

The time between conception and first documentation of pregnancy in automated databases may vary greatly among women because of several factors, including variations in the symptoms of early pregnancy, health care insurance status, pregnancy planning, and other socioeconomic characteristics. As a result, correct classification of first-trimester exposures with algorithms based on pregnancy indicators is difficult. Even when we used the date of the first pregnancy test, which takes place closer to the actual date of conception but is often self-administered at home and would not be recorded in most databases, the accuracy of this algorithm was suboptimal. When we used the date of the first prenatal visit, a more realistic estimate of the first pregnancy indicator in automated databases, the sensitivity of this algorithm was less than 60 percent. Moreover, the algorithm excludes by definition women who begin prenatal care late. Under various scenarios, such women might be more or less likely to use medications during the first trimester. Therefore, even if the algorithm were reliable in terms of timing of exposure among the included women, use of drugs by these women might not represent that in the general population.

The 70-day interval used in the pregnancy-indicator algorithm was shorter than the 90-day length commonly used to define trimesters of pregnancy, resulting in a reduced capacity to capture actual antiinfective use during the first trimester. However, the pregnancy-indicator algorithm still performed relatively poorly even when we used a 90-day definition. In our study, the first prenatal visit was, on
average, 36 days after the estimated date of conception, implying that the pregnancy-indicator algorithm missed the first weeks of the first trimester and incorrectly included part of the second trimester. Thus, algorithms based on first pregnancy indicators recorded in automated databases might poorly represent drug use during the first trimester of pregnancy.

Factors that affect the accuracy of algorithms

The extent of exposure misclassification depends not only on the algorithms used to categorize different periods of pregnancy but also on the patterns of drug use and the length of the period of interest. In this paper, we have presented data on antiinfectives, which are typically used for 7–10 days, and have focused on an exposure window of 3 months. Longer or repeated treatment episodes (e.g., antidepressants, antiepileptics) would tend to be associated with higher sensitivities; in the extreme, a drug used continuously during pregnancy would result in 100 percent sensitivity even if there were no overlap between the true gestational period and that defined by the algorithm. Similarly, the longer the period of interest, the easier it will be to correctly classify drug use. For example, if a given algorithm misses the date of conception by 1 week and the mother used the drug during the fifth week of gestation, the algorithm might misclassify exposure if we define weekly periods, but it would accurately classify this woman as exposed during the first trimester.

Potential impact of misclassification on relative risk estimates

For a binary exposure, if we assume no outcome misclassification and nondifferential misclassification of the exposure, the bias in the effect estimate due to imperfect sensitivity and specificity will be toward the null (9). Under differential misclassification of the exposure, or polytomous exposure, the direction and magnitude of bias are more difficult to predict. Data on medication use are often collected prospectively in automated databases; therefore, misclassification of exposure is usually considered nondifferential. However, algorithms can introduce differential misclassification of exposure if their ability to time pregnancy periods is worse for preterm births. Using an algorithm that is less accurate for births at younger gestational ages is of concern because there are many circumstances in which one might expect birth outcomes to be directly or indirectly associated with time of delivery. Estimates of the effects of drug exposure during pregnancy are more likely to be valid if pregnancy-related data such as LMP and gestational age are collected through data linkage with external sources (e.g., birth certificates).

Other limitations of automated databases

Automated databases have become an attractive data source for research related to periconceptual medication use. However, in addition to the problems identified above, these databases usually lack information on important pregnancy-related (e.g., parity) and lifestyle (e.g., smoking, alcohol intake, dietary folate) factors and underlying indications for the drugs; they might also have incomplete link-age of mother-baby pairs or incomplete follow-up of gestation (e.g., spontaneous abortions might be missed). Furthermore, they are generally constrained by the absence of over-the-counter drug information, and the prescription or dispensing data recorded in the automated databases do not necessarily reflect actual use. Some of these disadvantages can be diminished by restricting the studies to prescription medications and validating automated data through review of medical records or interviews of a random subset of the study cohort or physicians.

Limitations of the BDS

The data used for the current study have several strengths and limitations as well. The fact that BDS has detailed information on timing of pregnancy makes it an appropriate database to define a gold standard for timing of gestational periods. Although BDS data are collected retrospectively and inaccurate recall may occur, careful approaches are used to maximize completeness and accuracy of recall and reduce this bias (6). Moreover, the aim of this study was to compare not the accuracy of drug exposure ascertainment but the ability to pinpoint periods of exposure. That is, given that a woman used an antifungicide, how accurately can an algorithm discriminate the gestational period of exposure? Therefore, even if BDS participants underreported use of antibiotics, the sensitivities and specificities estimated in our analyses would be valid unless reported and unreported exposures represented very different patterns of antibiotic use.

Generalizability of findings

Although the methods used by Andrade et al. (4) and Hardy et al. (7) were based on two of the most commonly used computer algorithms, others may use different approaches, and our results may not be readily generalizable to other studies. For example, a variation in the delivery-date algorithm can also be found in recent literature (10). However, identification of the onset of gestation is probably the most salient difference between the two approaches we studied, and our findings should provide insight into the performance of these commonly applied algorithms. Similarly, our results would not be generalizable to other medications with different patterns of use, as discussed above.

Conclusions

In the absence of information on timing of conception, using the date of delivery as the reference date to identify antiinfective exposure during the first trimester among women with term pregnancies appeared to be a reasonable method. However, the accuracy of this method was diminished for women with conditions associated with shorter gestations, such as pregnancy-related complications or fetal abnormalities. Therefore, the question remains whether databases without adequate information on timing of
pregnancy can be used to provide valid information on the risk and safety of drugs during pregnancy.

ACKNOWLEDGMENTS

Dr. Hernández-Díaz was partially supported by grant R01 HD046595 from the National Institute of Child Health and Human Development.

The authors thank Dawn Jacobs, Fiona Rice, Rita Krolak, Kathleen Sheehan, Karen Bennett Mark, Clare Coughlin, Nastia Dynkin, Nancy Rodriguez-Sheridan, and Meghan Malone-Moses for their assistance with data collection and computer programming; the staff of the Massachusetts Department of Public Health Center for Birth Defects Research and Prevention; and all the mothers who participated in the study.

Conflict of interest: none declared.

REFERENCES


