Misleading meta-analyses of observational studies may generate unjustified alarms: The case of medications for nausea and vomiting in pregnancy

Annalisa Biffi\textsuperscript{ab} Federico Rea\textsuperscript{ab} Anna Locatelli\textsuperscript{cd} Irene Cetin\textsuperscript{ef} Amelia Filippelli\textsuperscript{ag} Giovanni Corrao\textsuperscript{ab}

a National Centre of Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
b Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
c Department of Medicine and Surgery, University of Milano-Bicocca, Milano, Italy
d Department of Obstetrics and Gynecology, Ospedale San Gerardo, Monza, Italy
e Scienze biomediche e cliniche 'Luigi Sacco', University of Milano, Milano, Italy
f ASST Fatebenefratelli Sacco, Milano, Italy
g Department of Medicine and Surgery, University of Salerno, Salerno, Italy

ABSTRACT

Objectives

Because observational studies often use imperfect measurements, results are prone to misclassification errors. We used as a motivating example the possible teratogenic risks of antiemetic agents in pregnancy since a large observational study recently showed that first-trimester exposure to doxylamine-pyridoxine was associated with significantly increased risk of congenital malformations as a whole, as well as central nervous system defects, and previous observational studies did not show such associations. A meta-analysis on this issue was carried out with the aim to illustrate how differential exposure and outcome misclassifications may lead to uncertain conclusions.

Methods

Medline, searched to October 2019 for full text papers in English. Summary Odds Ratios (ORs) with confidence intervals (CIs) were calculated using random-effect models. Probabilistic sensitivity analyses were performed for evaluating the extension of differential misclassification required to account for the exposure-outcome association.

Results
Summary ORs were 1.02 (95% CI, 0.92–1.15), 0.99 (0.82–1.19) and 1.25 (1.08–1.44) for overall congenital, cardiocirculatory, and central nervous system malformations respectively. By assuming exposure and outcome bias factor respectively of 0.95 (i.e., newborns with congenital defects had exposure specificity 5% lower than healthy newborns) and 1.12 (i.e., exposed newborns had outcome sensitivity 12% higher than unexposed newborns), summary OR of central nervous system defects became 1.13 (95% CI, 0.99–1.29) and 1.17 (95% CI, 0.99–1.38).

**Conclusion**

Observational investigations and meta-analyses of observational studies need cautious interpretations. Their susceptibility to several, often sneaky, sources of bias should be carefully evaluated.

**KEYWORDS**

Doxylamine-pyridoxine
Pregnancy
Congenital malformations
Exposure misclassification
Outcome misclassification
Probabilistic sensitivity analysis
Meta-analysis

**1. INTRODUCTION**

The involvement of pregnant women in randomized controlled trials (RCTs) is a complex issue [1]. The concept of neonatal safety as the main determinant of study design, and the pregnant patient’s obligation towards the fetal patient [2], represent challenging ethical issues. Implications regard the general lack of RCTs with pregnant participants even for common therapies, such as those acting as antiemetic, and the extensive use of evidence-based observational primary studies and their systematic reviews and meta-analyses [3]. However, because it is common for each observational study the use of imperfect measurements as a surrogate for the true status, results from meta-analyses of observational studies are prone to misclassification errors [4].

We faced this issue through a motivating example of possible teratogenic risks of antiemetic agents in pregnancy. In particular, our attention focused on exposure to doxylamine-pyridoxine, the molecule acting as active principle in the most dispensed antiemetic drug currently used for the treatment of nausea and vomiting in pregnancy. The drug is known with the trade name Bendectin (United States), Debendox (United Kingdom), Diclectin (Canada) and Lenotan (other countries). The long history of Bendectin includes marketing interruption in the years’ 80 (for manufacturer decision due to concerns on fetal safety [5]), and the following re-approval in the’ 90 (due to FDA
statement on lack of evidence of its effect on birth defects [6,7]). Nowadays, exposure to doxylamine-pyridoxine in pregnancy (or to the combination of doxylamine and pyridoxine) is classified as safe [8], and its use is recommended as first-line therapy of nausea and vomiting in pregnancy by the Society of Obstetricians and Gynecologists of Canada [9], the American College of Gynecology and Obstetrics [10] and organizations such as Mother to Baby (formerly known as Organization of Teratogen Information Services [11]). For this reason, the results of a large observational study recently published, which provides new evidence that exposure to doxylamine-pyridoxine during the first trimester of pregnancy was associated with a significantly increased risk of overall and specific major congenital malformations [12], raise concerns among physicians and women, and alarms for regulatory agencies and manufacturers.

With these premises, a meta-analysis of observational studies investigating the relationship between maternal exposure to doxylamine-pyridoxine during the first trimester of pregnancy and newborn congenital malformation was performed. Focusing on sources of uncertainty, particularly exposure and outcome differential misclassification, was of particular interest in this study. In particular, we carried out a probabilistic sensitivity analysis to explore the extension of misclassification required to fully account for the observed exposure-outcome association. The Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines were followed for designing and reporting the current investigation (Supplementary Table S1) [13].

2. METHODS

2.1. Search strategy and study selection
We performed a PubMed search for observational studies published up to 8 October 2019 that investigated the association between exposure to doxylamine-pyridoxine during the first trimester of pregnancy and the development of neonatal malformations overall considered, or affecting specific sites such as cardiocirculatory and of central nervous system (CNS) defects. Search strategy contained keywords and/or corresponding MeSH terms related to pregnancy, antiemetic and malformations. Full details on the applied strategy are showed in Supplementary material (Table S2). We excluded original papers that i) were not published in English language, ii) did not report original findings (i.e., letters, case report, systematic review and meta-analysis), iii) did not consider doxylamine-pyridoxine use in the first trimester of pregnancy, iv) evaluated another class of antiemetics than that of interest, v) evaluated other outcomes than those of interest. Moreover, a
hand-checking search was performed in order to identify additional relevant studies. The flow
diagram for the selection of the studies included in the meta-analysis is shown in Supplementary
material (Fig. S1) and the reason of exclusion for articles reviewed by full-text are reported in the
Supplementary Table S3. Screening of articles was independently conducted by two reviewers (AB
and FR) and agreement between them was assessed using the Cohen Kappa statistic. Because the
value of the statistic was 0.847, an adequate inter-rater agreement was observed [14].

2.2. Data collection
For each of the twenty selected original investigations [12,[15], [16], [17], [18], [19], [20], [21],
[22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]], we extracted details on
publication year, country where the study was conducted, characteristics of the investigated
malformations (e.g., type and number), and odds ratio (OR) with 95 % confidence interval (CI). In
addition, information about the method for data collecting (self-report or pharmacy record) was
considered. Chronological summary of the included studies, and their main characteristics, is shown
in Table 1. Among the 20 included papers, 14 investigated congenital malformations as a whole
(including 380,080 pregnancies, 24,190 malformations and 62,176 doxylamine-pyridoxine users),
12 cardiocirculatory malformations (including 298,478 pregnancies, 6226 malformations and
59,640 doxylamine-pyridoxine users) and 11 CNS defects (including 339,756 pregnancies, 2763
malformations and 56,338 doxylamine-pyridoxine users).

2.3. Statistical analysis
The measure of interest was the summary OR that evaluated the relationship between exposure to
doxylamine-pyridoxine in the first trimester of pregnancy and congenital malformations,
considering mothers who did not use doxylamine-pyridoxine as reference. Where possible, we
pooled adjusted estimates from the original studies, raw data and computed unadjusted summary
OR were considered otherwise. Heterogeneity of between study-specific estimates was tested using
chi-square statistics [34] and measured with the I2 index (a percentage measure of variation across
studies caused by heterogeneity) [35]. We pooled the original estimates by fitting the DerSimonian
and Laird random-effects model [36]. Publication bias was investigated through funnel plot visual
analysis and the Egger’s test [37].

2.4. Quality assessment
The quality of the eligible studies was evaluated according to the Newcastle-Ottawa Scale (NOS)
for assessing the quality of nonrandomized study in meta-analysis [38]. High-quality studies were
defined as a score 6 or more of 9 total points. Moreover, the certainty of the evidence was assessed for each outcome using the GRADE approach, by considering the evidence from observational data [39]. The certainty of the pooled studies was downgraded using the risk of bias, imprecision, inconsistency, indirectness, and publication bias; otherwise there was an upgrading with a large effect, a dose–response effect, or absence of minimal residual confounding.

2.5. Accounting for uncertainty

Two sources of differential misclassification were envisaged and taken into account. As far as the use of antiemetic agents during pregnancy, we reasoned around different sources used by the included primary studies. Bérard et al. used the RAMQ (Regie de l’Assurance Maladie du Quebec) pharmaceutical service database including information from pharmacists’ claims for dispensed medications reimbursed by the program [40]. Exposure of interest was defined as having filled at least one prescription for doxylamine-pyridoxine, based on the assumption that drug prescriptions correspond to drug consumption. Validation of RAMQ database provided reassuring results, having found 78 % and 98 % positive and negative predictive values (respectively PPV and NPV) by comparing refill records with maternal interviews [41]. However, noncompliance in the use of prescribed medication is a well-known concern that may be particularly frequent among pregnant women owing to their fear of embryo-feto-toxic side effects of drugs [42]. Potential for differential exposure misclassification is a reasonable issue to be considered in studies assessing teratogenic effects of medications [[43], [44], [45]]. Compared with women who delivered a healthy newborn (noncases), those who delivered a baby with congenital defect (cases) might more frequently avoid drug use once it has been withdrawn in pharmacy (i.e., false-positive exposure might more frequently occur). For example, women who assumed drugs without harmful events and had given birth to a healthy baby in a previous pregnancy may be more confident and adherent to drug prescription and may have less risk to deliver a baby with a congenital defect in the next pregnancy, compared to nulliparous women. In this way, specificity is assumed differential between cases (SpE1) and noncases (SpE0) so that their ratio must be considered the exposure bias factor of interest (biasE = SpE1/SpE0). The impact of such bias on ORs of primary studies investigating maternal use of antiemetic drugs through prescription records [12,24,29] was evaluated by means of a sensitivity analysis [46,47]. The latter involved in building different scenario, so that biasE was left to range between 0.900 until 1.025. To evaluate the impact of exposure differential misclassification, we subsequently recalculated the summary OR for each biasE imposed value.
As far as the detection of congenital malformations, we reasoned as follows. ICD-9 and ICD-10 codes of congenital malformations recorded in the RAMQ database were validated against medical patient charts (PPV of at least 80 % and NPV of 93 %) [48]. However, as women who used antiemetic drugs could receive better medical attention and likely more accurate detection of newborn defects, the potential for differential outcome misclassification is a reasonable issue to be considered. In this way, sensitivity is assumed differential between exposed (SeY1) and non-exposed (SeY0), so that their ratio must be considered the outcome bias factor of interest (biasY = SeY1/SeY0). The impact of such bias on OR of each included primary study was again evaluated by means of a sensitivity analysis built by ranging biasY between values of 0.975 until 1.200. To evaluate the impact of outcome differential misclassification, we subsequently recalculated the summary OR for each biasY imposed value.

Monte Carlo simulations, each with 20,000 replications, were performed for calculating 95 % confidence interval of summary OR accounting for exposure and outcome differential misclassifications.

Full details on methods used for built probabilistic sensitivity analysis have been reported in Supplementary material.

3. RESULTS
3.1. Main studies results
Seven of the nineteen included studies had NOS scores of 6 or greater and were so assigned to the category of high-quality studies (Supplementary material, Table S4). The 19 studies published before the investigation by Bérard et al. were underpowered since the least OR detectable with power = 0.80 individually ranged from 1.4 [25,27] to >5.0 [33] for overall congenital malformations, from 1.5 [16] to >5.0 [29,30,33] for cardiocirculatory defects, and from 2.2 [20] to >5.0 [21,23,24,29] for congenital malformations of CNS (Table 1). In addition, only three of the 34 individually investigated associations (8.8 %) reached significance in ORs for cardiocirculatory malformations, one showing harmful effect [28], the other two protective action [16,17]. On the other hand, Béard & coll. found a significantly increased risk for overall congenital malformations (OR = 1.07) and for CNS defects (OR = 1.25) among doxylamine-pyridoxine users, which correspond to a number needed to harms (i.e., the number needed to generate an outcome due to the exposure) of 164 and 1,081, respectively. Because the authors investigated a cohort of 224,704
births (almost 59 % of the overall births included in the meta-analysis) of whom 45,598 resulted exposed (73 %), it is not surprising that the summary exposure - outcome association (henceforth E → Y) was strongly influenced by the last published study.

3.2. Exposure to doxylamine-pyridoxine and congenital malformations
Summarizing the available estimates of the 14 studies investigating congenital malformations as a whole, as well as cardiocirculatory defects, no statistical evidence that exposure to doxylamine-pyridoxine may affect these outcomes was observed (Fig. 1A and B). Conversely, summarizing the available estimates of the 11 studies investigating CNS defects, a significant effect was observed with summary OR equal to 1.25 (95 % CI, 1.08–1.44) (Fig. 1C). There was no evidence either for between-study heterogeneity for none of the investigated outcomes, nor for publication bias (Supplementary material, Fig. S2).

The certainty of these evidence was very low according to the GRADE approach (Supplementary material, Table S4).

3.3. Accounting for uncertainty
Small changes in exposure bias factor strongly affected the observed effect of doxylamine-pyridoxine on CNS malformations (Fig. 2A, left box). For example, summary ORs accounting for exposure differential misclassification became non-significant starting from exposure bias factor equal to 0.952. Thus, by assuming the latter bias factor value (Fig. 2A, right box), the adjusted summary OR became 1.13 (95 % CI, 0.99–1.29). On the other hand, the adjusted summary OR would become 1.00 if the exposure bias factor was 0.930.

Summary ORs accounting for outcome differential misclassification (Fig. 2B, left box) became non-significant starting from values of the bias factor equal to 1.124. Thus, by assuming the latter bias factor value (Fig. 2B, right box), the adjusted summary OR became 1.17 (95 % CI, 0.99–1.38).

4. DISCUSSION
It has been reported that around 60 % of women experience nausea and vomiting during the first trimester of their pregnancy [49], and that most of them are treated with drugs containing doxylamine-pyridoxine. It is therefore justified that the recent reporting of increased risk of neonatal congenital malformations associated with use of doxylamine-pyridoxine during the first
trimester of pregnancy [12] raised concerns. At least three reasons explain why the findings by Bérard et al. surprised everyone. First, although some plausible mechanisms involved in the role of doxylamine in the etiology of malformations have been proposed [23,50,51], none of the animal studies resulted in teratogenesis at exposures equivalent to or substantially above the human therapeutic dose of Bendectin [5]. Second, during the time frame when Bendectin was withdrawn from USA market, there was not observed concomitant reduction in those malformations alleged to be associated with Bendectin exposure [6,52]. More, the most recent Eurocat data regarding the prevalence of neural tube defects (NTD) in several European Countries did not show an increase (but instead a progressive reduction was reported) of detected NTD, especially in Spain where doxylamine-pyridoxine continued to be used in the last fifty years [53]. Finally, the massive amount of data so far published does not support a consistent statistical association between its use in pregnancy and a particular syndrome or group of malformations [54].

Our study investigated the potential of some common sources of error and misinterpretation of meta-analyses of observational studies, some of which have not received sufficient attention from the literature (Table 2).

First of all, we reasoned around some drawbacks common to studies investigating drug use during pregnancy by means of administrative drug registries, as for the study by Bérard et al. [12], as well as for other two previous investigations [24,29]. It is well known that, owing to their fear of embryo-feto-toxic side effects, some pregnant women do not use the prescribed and pharmacy refilled drugs, so generating a certain number of women incorrectly classified as exposed (false positive) [40]. It is reasonable that such misclassification is differential, i.e., it more affects women who have given birth to a newborn with congenital defects (cases), than other women (noncases). We observed that, even assuming very small differential between cases and noncases, the observed estimate results strongly biased. For example, if cases did not use the prescribed drug with a frequency of 4.8 % higher than noncases, then we should not reject the null hypothesis of exposure → outcome independence. Similarly, detection of congenital defects may differentially affect women who use, and those who do not use, antiemetic agents. For example, if the sensitivity of detecting CNS defects was 12 % higher among users than no users, than the observed exposure → outcome association should not be rejected. It should be emphasized the compatibility of such situations with the positive and negative predictive values reported by Bérard et al.
Of course, other common sources of bias may affect the considered estimates. For example, a bias away the null should be generated whether (i) over-the-counter use of drugs containing doxylamine-pyridoxine mainly concerned noncases than women who delivered a newborn with congenital defects (another source of differential misclassification of the exposure), (ii) use of doxylamine-pyridoxine was negatively correlated with supplementary use of folic acid before and during pregnancy, or positively correlated with the use of drugs increasing the risk of neonatal defects (confounding); (iii) pregnancies characterized by nausea and vomiting are biologically different from those not exhibiting those symptoms, leading the former to a worse prognosis (another source of confounding).

In addition, only seven of the nineteen (37 %) included studies were classified as “high-quality studies”, i.e., they had a NOS score of 6 or greater. Furthermore, language, publication and reporting biases may have affected our findings. However, few studies were excluded because written in other languages than English. Furthermore, if the studies that found no statistically significant increase in the risk of malformations associated with doxylamine-pyridoxine use had been less published or disseminated, the inclusion of them in our analysis should make the exposure and outcome misclassifications able to move the summarized estimate towards the null “smaller” than those we observed (i.e., differences in specificity between cases and noncases, and in sensitivity between exposed and non-exposed could be lower).

Finally, although Bérard et al. reported an increased risk of spina bifida and musculoskeletal system defects, we choose to focus on CNS malformations for three reasons. First, closure of the neural tube occurs already in the very first weeks of pregnancy, when the request for antiemetic drugs is not plausible. Second, the mean gestational age of the first prescription was reported to be around 8.2 weeks (approximately 6.2 weeks after conception), while it is well known that neural tube closure is completed around 28 days after conception [55]. Finally, though significant, musculoskeletal system defects were much more weakly associated with doxylamine-pyridoxine exposure than those of CNS (being the reporting punctual estimates of OR 1.08 and 1.25).

5. CONCLUSIONS
The current paper, rather than generate new evidence about possible teratogenic risks of antiemetic agents in pregnancy, must be understood as an exercise aimed to alert patients, physicians, and decision-makers on the intrinsic uncertainty of meta-analyses of observational studies whose results
should not go beyond hypothesis generation rather than provoke unnecessary alarms and anxieties, and seriously interfere with good medical practice.

The alarming conclusions of the paper by Bérard et al. [12] should be resized, especially in consideration of its susceptibility to several sources of bias, other than the lack of consistent biological mechanisms explaining it. At the same time, however, we believe that a detailed scrutiny of all available databases by health authorities and scientific groups should be encouraged, and urgently carried out.

6. DISCLOSURES
This study was supported by grants from the Italian Ministry of Education, University and Research (MIUR) (‘Fondo d’Ateneo per la Ricerca’ portion, year 2019). The funding source had no role in the design of the study, the collection, analysis and interpretation of the data, or the decision to approve publication of the finished manuscript.

DECLARATION OF COMPETING INTEREST
The authors report no declarations of interest.
Table 1. Chronological summary of studies investigating the relationship between exposure to doxylamine-pyridoxine during the first trimester of pregnancy and congenital malformations.

<table>
<thead>
<tr>
<th>First author publication year, country [ref]</th>
<th>Type of malformations</th>
<th>Newborns with / without malformations among those who were Exposed</th>
<th>Unexposed</th>
<th>Odds ratio (95% confidence interval)</th>
<th>Least odds ratio detectable with power = 0.80</th>
<th>Method for investigating drug exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunde 1963, America [33]</td>
<td>Overall</td>
<td>11/2207</td>
<td>21/2197</td>
<td>0.52 (0.25–1.08)</td>
<td>&gt;5.00</td>
<td>Interview</td>
</tr>
<tr>
<td>GPRG 1963, UK [32]</td>
<td>Overall</td>
<td>2/2216</td>
<td>1/2217</td>
<td>2.00 (0.18–22.08)</td>
<td>&gt;5.00</td>
<td>Interview</td>
</tr>
<tr>
<td>Milkovich 1976, America [31]</td>
<td>Overall</td>
<td>5/623</td>
<td>148/9429</td>
<td>0.51 (0.21–1.25)</td>
<td>2.17</td>
<td>Prescription records</td>
</tr>
<tr>
<td>Newman 1977, Australia [20]</td>
<td>Cardiocirculatory</td>
<td>6/1186</td>
<td>70/6671</td>
<td>0.48 (0.21–1.11)</td>
<td>2.12</td>
<td>Interview</td>
</tr>
<tr>
<td>Smithells 1978, UK [29]</td>
<td>Cardiocirculatory</td>
<td>1/1191</td>
<td>8/6733</td>
<td>0.71 (0.09–5.66)</td>
<td>&gt;5.00</td>
<td>Prescription records</td>
</tr>
<tr>
<td>Rothman 1979, America [28]</td>
<td>Cardiocirculatory</td>
<td>24/46</td>
<td>366/1208</td>
<td>1.72 (1.04–2.86)</td>
<td>2.13</td>
<td>Interview</td>
</tr>
<tr>
<td>Heinonen 1980, America [27]</td>
<td>Overall</td>
<td>63/937</td>
<td>3185/46,097</td>
<td>0.97 (0.75–1.26)</td>
<td>1.41</td>
<td>Not available</td>
</tr>
<tr>
<td>Fleming 1981,</td>
<td>Overall</td>
<td>7/993</td>
<td>259/49,023</td>
<td>1.33 (0.63–2.83)</td>
<td>2.62</td>
<td>Interview</td>
</tr>
</tbody>
</table>

Overall 8/475                               | 291/13,910             | 0.81 (0.40–1.63)                                              | 2.15     | Interview                          |
<table>
<thead>
<tr>
<th>First author publication year, country [ref]</th>
<th>Type of malformations</th>
<th>Newborns with / without malformations among those who were Exposed</th>
<th>Unexposed</th>
<th>Odds ratio (95% confidence interval)</th>
<th>Least odds ratio detectable with power = 0.80</th>
<th>Method for investigating drug exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotland [26]</td>
<td>Overall</td>
<td>78/1607</td>
<td>245/5526</td>
<td>1.09 (0.84–1.42)</td>
<td>1.43</td>
<td>Interview</td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>10/1675</td>
<td>41/5730</td>
<td>0.83 (0.42–1.67)</td>
<td>2.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central nervous system</td>
<td>5/1680</td>
<td>12/5759</td>
<td>1.43 (0.50–4.06)</td>
<td>3.52</td>
<td></td>
</tr>
<tr>
<td>Gibson 1981, Australia [25]</td>
<td>Overall</td>
<td>24/2231</td>
<td>56/2351</td>
<td>0.87 (0.54–1.41)</td>
<td>1.63</td>
<td>Prescription records</td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>5/2250</td>
<td>10/4572</td>
<td>1.02 (0.35–2.98)</td>
<td>3.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central nervous system</td>
<td>1/2254</td>
<td>2/4580</td>
<td>1.02 (0.09–11.21)</td>
<td>&gt;5.00</td>
<td></td>
</tr>
<tr>
<td>Jick 1981, America [24]</td>
<td>Overall</td>
<td>44/78</td>
<td>659/1634</td>
<td>1.40 (0.96–2.05)</td>
<td>1.75</td>
<td>Interview</td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>13/109</td>
<td>189/2104</td>
<td>1.33 (0.73–2.40)</td>
<td>2.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anencephaly/Spina bifida</td>
<td>2/120</td>
<td>14/2279</td>
<td>2.71 (0.61–12.07)</td>
<td>&gt;5.00</td>
<td></td>
</tr>
<tr>
<td>Eskenazi 1982, America [23]</td>
<td>Major malformations</td>
<td>20/931</td>
<td>175/11,192</td>
<td>1.37 (0.86–2.19)</td>
<td>1.96</td>
<td>Interview</td>
</tr>
<tr>
<td>Michaelis 1983, Germany [22]</td>
<td>Anencephaly/Spina bifida</td>
<td>2/5</td>
<td>7/22</td>
<td>1.26 (0.20–7.97)</td>
<td>&gt;5.00</td>
<td>Interview</td>
</tr>
<tr>
<td>Hearey 1984, America [21]</td>
<td>Central nervous system</td>
<td>31/30</td>
<td>733/734</td>
<td>1.03 (0.62–1.73)</td>
<td>2.21</td>
<td>Interview</td>
</tr>
<tr>
<td>Winship 1984, UK [20]</td>
<td>Overall</td>
<td>30/1550</td>
<td>75/4854</td>
<td>1.25 (0.82–1.92)</td>
<td>1.82</td>
<td>Prescription records</td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>6/1574</td>
<td>12/4917</td>
<td>1.56 (0.59–4.17)</td>
<td>4.02</td>
<td></td>
</tr>
<tr>
<td>First author publication year, country, ref</td>
<td>Type of malformations</td>
<td>Newborns with / without malformations among those who were</td>
<td>Odds ratio (95% confidence interval)</td>
<td>Least odds ratio detectable with power = 0.80</td>
<td>Method for investigating drug exposure</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Zierler 1985, America [18]</td>
<td>Cardiocirculatory</td>
<td>52/121, 240/607</td>
<td>1.09 (0.76–1.55)</td>
<td>1.66</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>51/2720, 520/28,273</td>
<td>0.70 (0.50–1.20)</td>
<td>1.46</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>16/2755, 299/28,494</td>
<td>0.55 (0.33 to 0.92)</td>
<td>1.61</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central nervous system</td>
<td>7/2764, 61/28,732</td>
<td>1.19 (0.55–2.61)</td>
<td>2.51</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td>Shiono 1989, America [17]</td>
<td>Cardiocirculatory</td>
<td>67/287, 334/998</td>
<td>0.70 (0.52 to 0.94)</td>
<td>1.46</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td>Boneva 1999, America [16]</td>
<td>Cardiocirculatory</td>
<td>6/19, 705/3997</td>
<td>1.84 (0.71–4.78)</td>
<td>3.68</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td>Anderka 2012, America [15]</td>
<td>Central nervous system</td>
<td>3945/41,653, 14,402/164,704</td>
<td>1.07 (1.03–1.11)</td>
<td>1.06</td>
<td>Prescription records</td>
<td></td>
</tr>
<tr>
<td>Bérard 2019, America [12]</td>
<td>Overall</td>
<td>972/44,626, 3552/175,554</td>
<td>1.02 (0.95–1.11)</td>
<td>1.15</td>
<td>Prescription records</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cardiocirculatory</td>
<td>225/45,373, 665/178,441</td>
<td>1.25 (1.06–1.47)</td>
<td>1.25</td>
<td>Prescription records</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Selected weaknesses and cautions of meta-analyses from observational studies.

<table>
<thead>
<tr>
<th>Weaknesses</th>
<th>Solutions and cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potential for differential exposure misclassification</strong></td>
<td></td>
</tr>
<tr>
<td>Can we reasonably assume that drug prescriptions correspond to drug use alike for cases and noncases?</td>
<td>Women who delivered an infant with congenital defect may have avoided more frequently the use of antiemetic medications, although the drug was withdrawn from pharmacy</td>
</tr>
<tr>
<td>Do data source on drug exposure alike complete for cases and noncases?</td>
<td>Compared with women who delivered an infant with congenital defect, noncases may more often had received over-the-counter antiemetic medications</td>
</tr>
<tr>
<td><strong>Potential for differential outcome misclassification</strong></td>
<td></td>
</tr>
<tr>
<td>Is diagnostic carefulness of congenital defects unaffected from drug exposure?</td>
<td>Women who used antiemetic medications may have received better medical care, the latter surrogating more accurate detection of newborn defects</td>
</tr>
<tr>
<td><strong>Potential for excessive influence of a single study</strong></td>
<td></td>
</tr>
<tr>
<td>Is overall pooled estimate influenced by a single study?</td>
<td>Summarized estimates may be strongly affected by a single study with sample size very larger than those of other investigations</td>
</tr>
<tr>
<td><strong>Lack of a clear causal hypothesis</strong></td>
<td></td>
</tr>
<tr>
<td>Is the tested hypotheses supported by clear biological mechanisms?</td>
<td>Closure of neural tube in the early pregnancy cannot be causally associated with antiemetic drug use</td>
</tr>
</tbody>
</table>
Fig. 1. Forest plots of study-specific and summary odds ratios for a) overall congenital malformations, or b) cardiocirculatory malformations and c) central nervous system defects.

Footnote. Squares represent study-specific odds ratio estimates (size of the square reflects the study-specific statistical weight, that is, the inverse of the variance); horizontal lines represent 95% CIs; diamonds represent summary odds ratio estimates with corresponding 95% CIs; p values are from testing for heterogeneity between study-specific estimates.
**Fig. 2.**

**a)** Influence of exposure bias factor on the summary odds ratio of nervous system defects associated with exposure to doxylamine-pyridoxine during the first trimester of pregnancy (left box) and corresponding forest plot of study-specific and summary adjusted odds ratios (right box).

Footnote. The bias factor is the assumed ratio between exposure specificity among cases (i.e., women who delivered a newborn with nervous system malformation) and noncases (i.e., women who delivered a healthy newborn). Data were simulated by varying bias factors between 0.900 and 1.025. Monte Carlo simulations with 20,000 replications were performed for calculating 95% confidence interval of summary OR accounting for exposure differential misclassification. Adjusted forest plot was built according with the minimum exposure bias factor value which makes the summary OR no significant (i.e., bias = 0.952)

**b)** Influence of outcome bias factor on the summary odds ratio of nervous system defects associated with exposure to doxylamine-pyridoxine during the first trimester of pregnancy (left box) and corresponding forest plot of study-specific and summary adjusted odds ratios (right box).

Footnote. The bias factor is the assumed ratio between outcome sensitivity among users (i.e., women who used antiemetic drugs during the first trimester of pregnancy) and no users. Data were simulated by varying bias factors between 0.975 and 1.200. Monte Carlo simulations with 20,000 replications were performed for calculating 95% confidence interval of summary OR accounting for outcome differential misclassification. Adjusted forest plot was built according with the minimum exposure bias factor value which makes the summary OR no significant (i.e., bias = 1.124)
REFERENCES

Addiction, 106 (2011), pp. 1355-1362

[2] L.B. McCullough, J.H. Coverdale, F.A. Chervenak *A comprehensive ethical framework for responsibly designing and conducting pharmacologic research that involves pregnant women*  


[5] R.L. Brent *Bendectin: review of the medical literature of a comprehensively studied human nonteratogen and the most prevalent tortogen-litigen*  

[6] Congenital malformation surveillance bulletin  
Teratology, 48 (1993), pp. 545-709


Pharmacology, 100 (2017), pp. 161-171

[9] M.Y. Arsenault, C.A. Lane *The management of nausea and vomiting in pregnancy*  
(Society of Obstetricians and Gynaecologists Canada: clinical Practice Guidelines, no. 120)  

[10] APGONausea and Vomiting of Pregnancy. APGO Educational Series on Women’s Health Issues  


[14] J.R. Landis, G.G. Koch The measurement of observer agreement for categorical data
Biometrics, 33 (1) (1977), pp. 159-174


Am. J. Epidemiol., 149 (1999), pp. 717-725

Teratology, 40 (1989), pp. 151-155


Am. J. Epidemiol., 120 (1984), pp. 559-564

Teratology, 27 (1983), pp. 57-64
[23] B. Eskenazi, M.B. Bracken. **Bendectin (Debendox) as a risk factor for pyloric stenosis**

JAMA, 246 (1981), pp. 343-346


[27] O.P. Heinonen. **Legal testimony: David Mekdecı v. Merrell National Laboratories; United States District Court, Middle District of Florida**

Am. J. Epidemiol., 109 (1979), pp. 433-439

Teratology, 17 (1978), pp. 31-35


[31] L. Milkovich, B.J. van den Berg. **An evaluation of the teratogenicity of certain antinauseant drugs**
Am. J. Obst. Gynecol., 125 (1976), pp. 244-248

Practitioner, 191 (1963), pp. 775-780

[33] C.A. Bunde, D.M. Bowles. **A technique for controlled survey of case records**

[34] W.G. Cochran. **The combination of estimates from different experiments**
Biometrics, 10 (1954), pp. 101-129

BMJ, 327 (2003), pp. 557-560

BMJ, 315 (1997), pp. 629-634

[38] A. Stang. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses.


[41] J.P. Zhao, O. Sheehy, J. Gorgui, et al. Can we rely on pharmacy claims databases to ascertain maternal use of medications during pregnancy?
Birth Defects Res., 109 (2017), Article 423e31

Epidemiology, 12 (2001), pp. 497-501

PLoS One, 7 (2012), Article e46626


Int. J. Epidemiol., 34 (2005), pp. 1370-1376

[48] L. Blais, A. Berard, F.Z. Kettani, et al. *Validity of congenital malformation diagnostic codes recorded in Quebec’s administrative databases*  

[49] A. Seto, T. Einarson, G. Koren *Pregnancy outcome following first trimester exposure to antihistamines: meta-analysis*  

[50] W.E. Serafin, K.S. Bab *H1-receptor antagonists*  

[51] A.S. Tani *Antihistamines*  
K. Olson (Ed.), Poisoning & Drug Overdose, Appletin & Lange, Norwalk, CT (1998), pp. 78-79

[52] D.W.G. Harron, K. Griffiths, R.G. Shanks *Debendox and congenital malformations in Northern Ireland*  

[53] *Eurocat Data: Prevalence Charts and Tables*  
European Commission (2019)  

[54] R.L. Brent *Method of evaluating alleged human teratogens (editorial)*  
Teratology, 17 (1978), p. 83

[55] E. Cardonick, A. Iacobucci *Use of chemotherapy during human pregnancy*  