Pregnancy and Delivery Outcomes Following Benzodiazepine Exposure: A Systematic Review and Meta-analysis

Résultats de la grossesse et de l'accouchement suite à une exposition aux benzodiazépines : une revue systématique et une méta-analyse

Sophie Grigoriadis, MD, MA, PhD, FRCPC[1,2,3,4], Lisa Graves, MD[5],
Miki Peer, PhD[2], Lana Mamisashvili, MSW[2], Myuri Ruthirakuhan, PhD Student[6],
Parco Chan, MSc Student[6], Mirna Hennawy, BSc[3], Supriya Parikh, BA[3],
Simone Natalie Vigod, MD, MSc, FRCPC[4,7], Cindy-Lee Dennis, PhD[8],
Meir Steiner, MD, FRCPC[9], Cara Brown, MD, FRCPC[4],
Amy Cheung, MSc, MD, FRCPC[1,3,4], Hiltrud Dawson, RN[10], Neil Rector, PhD[1,3,4],
Melanie Guenette, MSc[2,11], and Margaret Richter, MD, FRCPC[1,3,4]

Abstract

Objective: Understanding the effects of benzodiazepines (BZDs) on maternal/fetal health remains incomplete despite their frequent use. This article quantifies the effects of antenatal BZD exposure on delivery outcomes.

Methods

Data Sources: Medline, PsycINFO, CINAHL, Embase, and the Cochrane Library were searched till June 30, 2018.

Study Selection: English-language cohort studies comparing antenatal BZD exposure to an unexposed group on any delivery outcome were eligible. In all, 23,909 records were screened, 56 studies were assessed, and 14 studies were included.

Data Extraction: Two reviewers independently assessed quality and extracted data. Estimates were pooled using random effects meta-analysis. Sub-analyses examined several potential moderators including timing of exposure.

Results: There were 9 outcomes with sufficient data for meta-analysis. Antenatal BZD exposure was significantly associated with increased risk of 6 outcomes initially: spontaneous abortion (pooled odds ratio = 1.86; 95% confidence interval [CI], 1.43 to 2.42), preterm birth (1.96; 95% CI, 1.25 to 3.08), low birth weight (2.24; 95% CI, 1.41 to 3.88), low Apgar score (2.19; 95% CI, 1.94 to 2.47), Neonatal Intensive Care Unit (NICU) admission (2.61; 95% CI, 1.64 to 4.14), and induced abortion (2.04; 95%

[1] Department of Psychiatry, Sunnybrook Health Sciences Centre, University of Toronto, Ontario, Canada
[2] Evaluative Clinical Sciences, Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, University of Toronto, Ontario, Canada
[3] Hurvitz Brain Sciences Program, Sunnybrook Research Institute, University of Toronto, Ontario, Canada
[4] Department of Psychiatry, University of Toronto, Ontario, Canada
[5] Department of Family and Community Medicine, Homer Stryker MD School of Medicine, Western Michigan University, Kalamazoo, MI, USA
[6] Department of Pharmacology and Toxicology, Sunnybrook Research Institute, University of Toronto, Ontario, Canada
[7] Department of Psychiatry, Women's College Hospital, University of Toronto, Ontario, Canada
[8] Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Ontario, Canada
[9] Department of Psychiatry & Behavioural Neurosciences, St. Joseph's Healthcare Hamilton, McMaster University, Ontario, Canada
[10] Health Nexus, Toronto, Ontario, Canada
[11] Division of Neurology, St. Michael's Hospital, University of Toronto, Ontario, Canada

Corresponding Author:
Sophie Grigoriadis, MD, MA, PhD, FRCPC, Women's Mood and Anxiety Clinic: Reproductive Transitions, Department of Psychiatry, Sunnybrook Health Sciences Centre, University of Toronto, FG 29, 2075 Bayview Ave., Toronto, Ontario, Canada M4N 3M5.
Email: sophie.grigoriadis@sunnybrook.ca

CI, 1.23 to 3.40). There was significant heterogeneity between studies for most outcomes without consistent moderators. Birth weight (mean difference [MD]: −151.35 g; 95% CI, −329.73 to 27.03), gestational age (−0.49 weeks; 95% CI, −1.18 to 0.19), and small for gestational age (SGA; 1.42; 95% CI, 1.00 to 2.01) did not show significant associations although after adjusting for publication bias, gestational age, and SGA became significant, totaling 8 significant outcomes.

Conclusions: Antenatal BZD exposure appears to be statistically associated with increased risk of several adverse perinatal outcomes. Although confounds cannot be ruled out, NICU admission does appear clinically relevant and consistent with the antidepressant literature.

Abrégé

Objectif : Comprendre les effets des benzodiazépines sur la santé de la mère et du fœtus demeure incomplet malgré l'usage fréquent de ces médicaments. Le présent article quantifie les effets de l'exposition prénatale aux benzodiazépines sur les résultats de l'accouchement.

Méthodes

Sources de données : Une recherche a été menée dans Medline, PsycINFO, CINAHL, Embase, et la Cochrane Library jusqu'au 30 juin 2018.

Sélection des études : Les études de cohortes en anglais comparant l'exposition prénatale aux benzodiazépines avec un groupe non exposé selon le résultat de l'accouchement étaient admissibles. La recherche a repéré 23 909 études, 56 ont été évaluées, et 14 études ont été incluses.

Extraction des données : Deux réviseurs ont évalué indépendamment la qualité et les données extraites. Les estimations ont été regroupées à l'aide d'une méta-analyse à effets aléatoires. Des sous-analyses ont examiné plusieurs modérateurs potentiels, notamment la durée d'exposition.

Résultats : Il y avait 9 résultats dont les données suffisaient pour la méta-analyse. L'exposition prénatale aux benzodiazépines était significativement associée à un risque accru de six résultats initialement: avortement spontané (rapport de cotes regroupées = 1.86; intervalle de confiance à 95% CI, 1.43 à 2.42), naissance prématurée (1.96; 1.25 à 3.08), faible poids à la naissance (2.24; 1.41 à 3.88), faible indice d'Agpar (2.19; 1.94 à 2.47), hospitalisation à l'unité de soins intensifs néonatals (USIN; 2.61; 1.64 à 4.14), et avortement provoqué (2.04; 1.23 à 3.40). Il y avait une hétérogénéité significative entre les études pour la majorité des résultats sans modérateur constant. Le poids à la naissance (différence moyenne (DM): −151.35 g; −329.73 à 27.03), l'âge gestationnel (DM = −0.49 semaines; −1.18 à 0.19), et nourrisson petit pour l'âge gestationnel (1.42; 1.00 à 2.01) n'indiquaient pas d'associations significatives bien qu'après ajustement pour biais de publication, l'âge gestationnel et nourrisson petit pour l'âge gestationnel devenaient significatifs, totalisant 8 résultats significatifs.

Conclusions : L'exposition prénatale aux benzodiazépines semble être statistiquement associée à un risque accru de plusieurs résultats périnatals indésirables. Bien que des facteurs de confusion ne puissent être exclus, l'hospitalisation à l'USIN semble cliniquement pertinente et compatible avec la littérature sur les antidépresseurs.

Introduction

While benzodiazepines (BZDs) are frequently prescribed for anxiety disorders, there are recommendations to limit their use. These drugs are generally not first-line treatment, but prescribing patterns suggest that they are used long term. Exposure during pregnancy can occur continuously or ad hoc, and research suggests the use of antianxiety medications during pregnancy has been increasing over time.[1-4] A large Swedish study (1996 to 2011) reported that sedatives/hypnotics were the third most commonly used central nervous system active drug in pregnancy, following antidepressants and opioids; 4.6% of infants were exposed to sedatives/hypnotics during the first trimester.[5] As such, knowledge about their safety during pregnancy is important, yet our understanding of their effects on maternal and fetal health remains incomplete.[1]

Reviews[6] and limited meta-analyses[7,8] have concluded that BZDs are not teratogenic. However, adverse perinatal outcomes have since been reported following antenatal exposure to BZDs including an increased risk for spontaneous abortion,[9,10] preterm birth,[5,11] low birth weight,[5,11] small for gestational age (SGA),[5] and low Apgar scores at 5 min.[5] Despite their ubiquitous use, to date, there has been no synthesis of the impact of antenatal BZD use with respect to diverse perinatal outcomes. Given the importance of this information for antenatal treatment decisions, this lack of synthesized data represents a gap with significant implications for clinical practice and maternal/infant health.

The aim of this study was to address this gap by systematically reviewing the data and meta-analyzing the available evidence to improve clinical decision-making. We pooled

data that were collected prospectively following antenatal BZD exposure in any trimester of pregnancy to examine the relationship between antenatal BZD exposure and various maternal and infant outcomes. The effects of study quality, timing of exposure, confounder adjustment (i.e., use of other psychoactive drugs, maternal psychiatric diagnosis), and geographic location were examined in moderator analyses.

Method

Search Strategy

This systematic literature review was guided by an Advisory Committee of key stakeholders and followed the guidelines of the Meta-Analyses and Systematic Reviews of Observational Studies.[12] Details of our methods have been previously described as this is one study in a program of research.[13] Two professional librarians with expertise in psychiatry and psychopharmacology conducted independent literature searches as part of a broader study of the impact of antenatal anxiety and antenatal use of anxiety-related medications. One librarian searched Ovid Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, CINAHL, Embase, and the Cochrane Library from their start date to June 30, 2018. Keywords included *anti-anxiety agents*, *anxiolytic*, *benzodiazepine*, *chlordiazepoxide*, *bromazepam*, *clonazepam*, *diazepam*, *lorazepam*, *nitrazepam*, *oxazepam*, *temazepam*, *pregnancy*, *prenatal or antenatal*, *preterm birth or premature delivery*, *low birth weight*, *birth weight*, *gestational age*, *Apgar*, *NICU*, *neonatal/infant outcomes*, *delivery outcomes*, among others (available upon request). The second librarian consulted with the review team and used an iterative process to develop and test the search strategy. This strategy was reviewed by another senior information specialist using the Peer Review for Electronic Search Strategies Checklist[14] prior to execution. Strategies utilized a combination of keywords (e.g., antenatal, anxiolytics) and controlled vocabulary (e.g., "pregnancy," "anxiety disorders," "anti-anxiety agents"), adjusting vocabulary and syntax as needed across databases (full search strategy available from the authors upon request). Results were limited to the English language, human studies, and nonopinion pieces. Reference lists from reviews, meta-analyses, and the final selection of included articles were also searched. The preferred reporting items for systematic review and meta-analyses (PRISMA) were followed.[15]

Inclusion and Exclusion Criteria

Studies with original, prospectively collected data were eligible if they included a BZD-exposed group and an unexposed comparison group. Acceptable assessment of BZD exposure included either filling a prescription for a BZD or notation of BZD use in maternal/clinical charts. Studies that examined outcomes in the mother or infant in the gestational, delivery, and/or postpartum period were accepted. For cases in which a sample was repeated in more than one publication, we chose the article that most closely addressed our research question or the one with larger sample size if that differed. Studies presenting only data aggregating BZDs with other psychotropic medication(s) were excluded. Case-control studies and one study where BZD was administered during an in-hospital stay for the treatment of severe hyperemesis gravidarum were excluded. Due to the volume of potentially eligible studies, abstracts, conference proceedings, and unpublished data were also excluded. Meta-analysis was conducted if there were at least 3 eligible studies.

Quality Assessment and Data Extraction

We used methods previously developed by our team for quality assessment and data extraction.[16] Two trained reviewers independently conducted quality assessments, and disagreements were resolved on a case-by-case basis by discussion with the study leads (SG, LG). The Systematic Assessment of Quality in Observational Research (SAQOR),[16] adapted from the Downs and Black[17] Checklist and the Newcastle–Ottawa Scale,[18] was used to assess article quality.

The studies were assessed on 19 aspects under the categories of sample (one example included if it was representative), control group (i.e., if one was included), quality of exposure/outcome measure (i.e., if the assessment was adequate), follow-up (i.e., if attrition rate was stated), and distorting influences/control for confounders. This last category, distorting influences, included whether the analyses controlled for confounders such as anxiety diagnosis, other psychotropic medications, and identification then control for additional confounders such as alcohol, smoking, or illicit substances. Assessments using the SAQOR criteria were combined with a modification of the Grading of Recommendations Assessment, Development, and Evaluation system,[19] to assign each article with a quality rating of high, moderate, low, or very low quality. This rating was further classified as either "above quality threshold" (high, moderate, and low ratings) or "below quality threshold" (very low rating).

The Strengthening the Reporting of Observational Studies in Epidemiology criteria[20] were used as the foundation for data extraction procedures. Extracted data included authors, year of publication, source country, details of study design, participants (sample, control, demographics, and clinical characteristics), inclusion/exclusion criteria, details of BZD use (timing and indication of use in pregnancy, BZD type, and dose), use of other psychotropics in pregnancy, outcomes and their assessment methods and definitions, statistical adjustment for confounders, and loss to follow-up. Adjusted estimates with their variances were extracted when available; when adjusted estimates were not provided in the published data, we calculated crude odds ratios (*OR*) or mean differences (*MDs*) and sample variances. Before calculating the *OR* for studies that included cells with a 0 count, we added 0.5 to these cells. A request for data analysis

clarification was sent to one study author and a response was received, allowing for inclusion of the study in our analyses. Data were extracted by one reviewer and checked by another and the primary author (SG); disagreements were resolved by discussion. Outcomes were as defined by the authors of the original publication.

Statistical Analyses

MD, *OR*, 95% confidence intervals (CIs), and heterogeneity were determined using Review Manager (RevMan) Version 5.3. Random effects models were used (as heterogeneity was expected across studies) for pooled estimates of the weighted MD (for continuous outcomes) and for *OR* (for binary outcomes). Relative risks were treated as *OR*s when events were rare (i.e., rate of <10%) and converted to *OR*s. Between-study heterogeneity was assessed with Cochrane Q and quantified with I^2, which represents the percentage of the between-study variance impacting quantitative results more than expected by chance, low ($I^2 = 25\%$), moderate ($I^2 = 50\%$), or high ($I^2 = 75\%$).[21] The intended quantitative and qualitative analyses for publication bias included Egger's test and visual inspection of funnel plots, respectively. In instances where publication bias was identified in analyses with 5 or more studies, the trim-and-fill method[22] was conducted to compute an adjusted *OR* or MD.

For outcomes with ≥ 4 studies, we examined potential sources of heterogeneity through subgroup analyses, regardless of the Q statistic as they were planned a priori. Subgroup analyses examined sources of heterogeneity: (a) study quality (above or below quality threshold based on the quality assessment), (b) timing of BZD exposure (first trimester, up to 22 weeks, any time in pregnancy), (c) adjustments (data adjusted/not adjusted for confounders), (d) use of other psychotropic medications (exclusion was not specified or assessed) or BZD monotherapy, (e) controlling for psychiatric diagnoses (adjusted data, excluded, not specified), and (f) geographic location (North America, Europe, other).

Patient Involvement

This study was conducted without patient involvement, although there was involvement at the beginning of the program of research and the advisory committee steered the team.

Results

We screened 23,909 records by title and abstract. Fifty-six articles were examined for inclusion, and 14 were included in the meta-analysis (Figure 1) with 9 outcomes including spontaneous abortion, preterm birth, birth weight, low birth weight, SGA, gestational age at birth, low Apgar score at 5 min, Neonatal Intensive Care Unit (NICU) admission, and induced abortion. Study characteristics are presented in Supplemental Table 1.[5,9-11,23-32] Most articles were above study

quality threshold and provided data on at least two different outcomes (13/14 and 13/14 studies, respectively).

Spontaneous abortion was significantly associated with BZD exposure (3,386 exposed and 1,204,620 unexposed) during pregnancy based on 5 pooled studies ($OR = 1.86$; 95% CI, 1.43 to 2.42, $P < 0.001$) although significant heterogeneity across studies was observed, accounting for 76% of the variance ($Q_4 = 16.83$, $P = 0.002$; Figure 2A). Subgroup analyses examining possible moderators found significant effects for all subgroups although 3 moderators were significant (Table 1). These were timing of exposure (with "anytime exposure" having the highest *OR*, based on 3 studies), exposure to other psychotropics ("not specified" having a higher *OR*, based on 4 studies), and controlling for psychiatric diagnoses ("not specified" having the highest *OR*, based on 3 studies).

Based on 6 pooled studies, use of BZDs (2,746 exposed and 1,327,706 unexposed) in pregnancy was significantly associated with *preterm birth* ($OR = 1.96$; 95% CI, 1.25 to 3.08, $P = 0.004$); however, heterogeneity across studies was significant and accounted for 83% of the variance ($Q_5 = 28.94$, $P < 0.001$; Figure 2B). Despite this, subgroup analyses did not identify a significant moderator (Table 1). Restricting analyses to the 5 studies that defined *preterm birth as <37 weeks* gestation also yielded a significant association with antenatal BZD exposure ($OR = 2.17$; 95% CI, 1.28 to 3.66, $P = 0.004$) and a similarly high level of heterogeneity across studies ($Q_4 = 26.89$, $P < 0.001$; $I^2 = 85\%$, potential moderator analyses also not significant). Interestingly antenatal BZD exposure (794 exposed and 3,410 unexposed) was not associated with *gestational age* (five studies, MD $= -0.49$ weeks; 95% CI, -1.18 to 0.19, $P = 0.16$). Significant heterogeneity across studies accounted for 91% of the variance ($Q_4 = 44.35$, $P < 0.001$), with none of the moderator analyses pooling to significance (Table 1).

Birth weight was not significantly lower in infants exposed to BZDs antenatally (794 exposed and 3,410 unexposed) based on 5 pooled studies (MD $= 151.35$ g; 95% CI, -329.73 to 27.03, $P = 0.10$); significant heterogeneity across studies accounted for 87% of the variance ($Q_4 = 31.90$, $P < 0.001$), with country being a significant moderator, although there were only, at most, 2 studies per group (Table 1). Pooling 6 studies, antenatal BZD exposure (2,725 exposed and 1,327,958 unexposed) was significantly associated with delivery of a *low birth weight* neonate ($OR = 2.34$; 95% CI, 1.41 to 3.88, $P = 0.001$). Significant heterogeneity, accounting for 85% of the variance, was observed across studies ($Q_4 = 32.40$, $P < 0.001$; Figure 2C). Subgroup analyses were all significant, indicating country as a significant moderator, with North American data pooling to a higher *OR*.

Three outcomes were based on pooling 4 studies. Both *low Apgar scores at 5 min* (2,385 exposed and 1,324,730 unexposed; $OR = 2.19$; 95% CI, 1.94 to 2.47, $P < 0.001$) and *NICU admission* (503 exposed and 3,811 unexposed; $OR = 2.61$; 95% CI, 1.64 to 4.14, $P < 0.001$) were significant
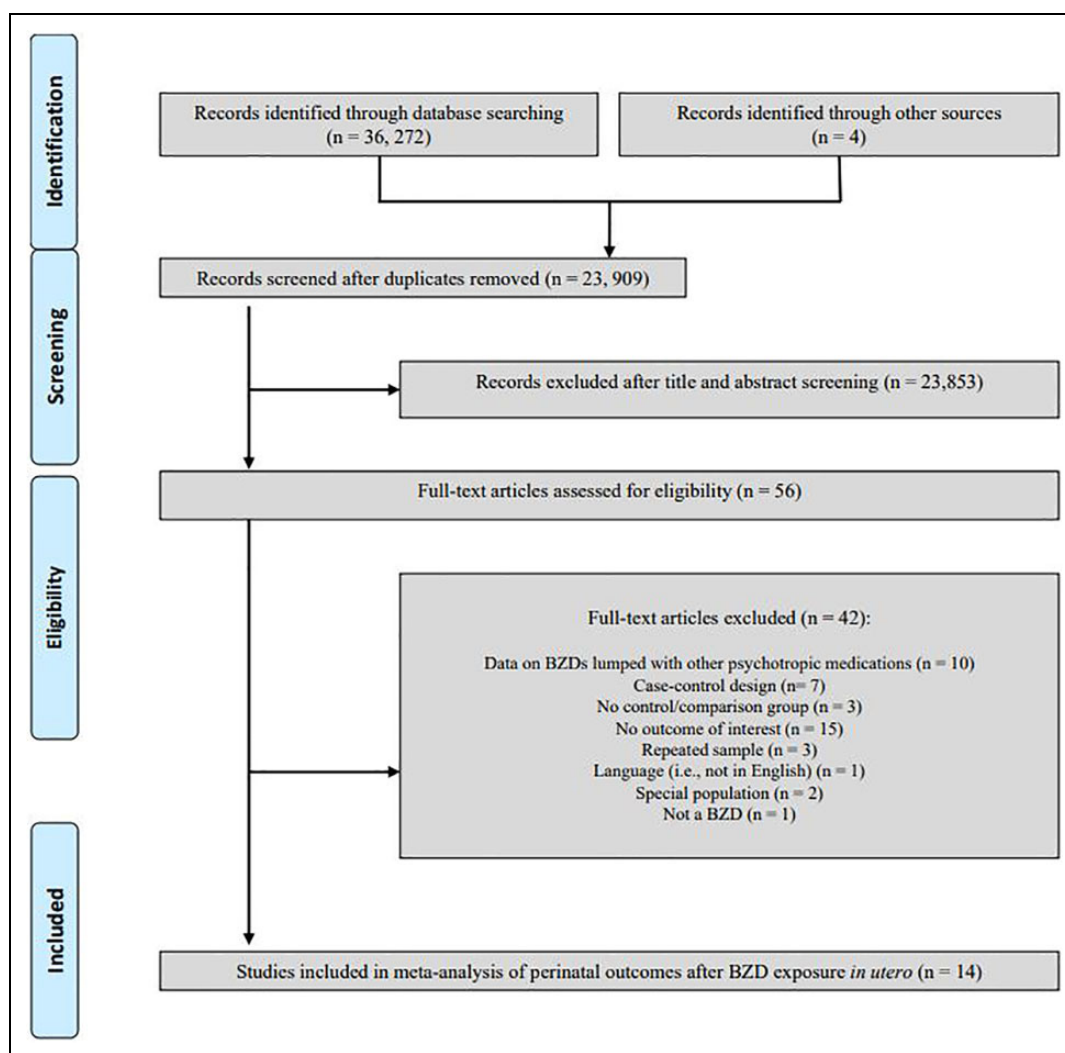
Figure 1. Flow diagram of study identification, selection, and reasons for exclusion in systematic review.

and neither showed significant heterogeneity across studies (respectively, $Q_3 = 2.92$, $P = 0.40$; $I^2 = 0.0\%$; Figure 2D, and $Q_3 = 6.70$, $P = 0.08$; $I^2 = 55.0\%$; Figure 2E). SGA was not significant (2,279 exposed and 1,324,081 unexposed; $OR = 1.34$; 95% CI, 0.97 to 1.86, $P = 0.08$), and heterogeneity was also not significant ($Q_2 = 5.54$, $P = 0.14$; $I^2 = 46.0\%$). The three studies that examined *induced abortion* pooled to a significant OR (2,981 exposed and 391,226 unexposed; $OR = 2.04$; 95% CI, 1.23 to 3.40, $P = 0.006$) with significant heterogeneity across studies ($Q_2 = 7.49$, $P = 0.02$; $I^2 = 73\%$).

Publication Bias

Publication bias was assessed visually with funnel plot and quantitatively with Egger's test when there were >5 studies. For *preterm birth* (6 studies) and *low birth weight* (6 studies), Egger's test did not indicate significant publication bias. Visual inspection of the preterm birth funnel plot based on <37 weeks (5 studies) did indicate slight bias. Applying the

trim-and-fill method resulted in a slightly higher and still significant OR ($OR = 2.51$, 95% CI, 1.49 to 4.24, $P < 0.001$). For *spontaneous abortion*, *birth weight*, and *gestational age*, visual inspection of the funnel plots indicated some publication bias. The revised OR using the trim-and-fill method for spontaneous abortion was slightly attenuated but remained significant ($OR = 1.43$; 95% CI, 1.10 to 1.86, $P < 0.001$). The effects of antenatal BZD exposure on birth weight remained insignificant, after imputing one missing study for potential publication bias (MD = -92.48 g; 95% CI, -267.09 to 82.13, $P = 0.30$). Interestingly, gestational age became significant after imputing two missing studies to account for publication bias (MD = -0.88 weeks; 95% CI, -1.60 to -0.16, $P = 0.02$). Visual inspection of the *Apgar at 5 min* outcome indicated slight publication bias, and the revised OR using trim-and-fill remained significant ($OR = 2.19$, 95% CI, 1.94 to 2.46, $P < 0.001$). The *NICU admission* outcome also appeared biased, but the revised OR using trim and fill, while slightly reduced, remained significant ($OR = 2.54$, 95% CI, 1.71 to 3.78, $P < 0.001$). *SGA* also appeared to

A

Study or Subgroup	Weight	Odds Ratio IV, Random, 95% CI
Ban 2012	30.1%	1.40 [1.26, 1.57]
Bech 2014	26.2%	1.91 [1.54, 2.37]
Milkovich 1974	22.7%	2.46 [1.82, 3.30]
Ornoy 1998	13.0%	1.93 [1.10, 3.40]
Pastusak 1994	8.0%	2.05 [0.91, 4.58]
Total (95% CI)	100.0%	1.86 [1.43, 2.42]

Heterogeneity: Tau^2 = 0.06; Chi^2 = 16.83, df = 4 (P = 0.002); I^2 = 76%
Test for overall effect: Z = 4.59 (P < 0. 00001)

B

Study or Subgroup	Weight	Odds Ratio IV, Random, 95% CI
Calderon-Margalit 2009	17.0%	6.79 [4.01, 11.50]
Freeman 2018	12.2%	1.31 [0.55, 3.12]
Kallen 2013	21.5%	1.62 [1.37, 1.91]
Kilic 2014	18.8%	1.60 [1.07, 2.40]
Sutter-Dallay 2015	16.3%	1.17 [0.66, 2.07]
Yonkers 2017	14.3%	1.98 [0.97, 4.04]
Total (95% CI)	100.0%	1.96 [1.25, 3.08]

Heterogeneity: Tau^2 = 0.24; Chi^2 = 28.94, df = 5 (P = 0.001); I^2 = 83%
Test for overall effect: Z = 2.92 (P = 0. 004)

C

Study or Subgroup	Weight	Odds Ratio IV, Random, 95% CI
Calderon-Margalit 2009	17.1%	7.43 [4.15, 13.30]
Freeman 2018	8.9%	2.33 [0.63, 8.62]
Kallen 2013	21.3%	1.40 [1.14, 1.72]
Kilic 2014	19.1%	1.83 [1.20, 2.80]
Sutter-Dallay 2015	18.8%	1.40 [0.89, 2.19]
Yonkers 2017	14.9%	3.41 [1.61, 7.22]
Total (95% CI)	100.0%	2.34 [1.41, 3.88]

Heterogeneity: Tau^2 = 0.30; Chi^2 = 32.40, df = 5 (P< 0.00001); I^2 = 85%
Test for overall effect: Z = 3.29 (P = 0. 001)

D

Study or Subgroup	Weight	Odds Ratio IV, Random, 95% CI
Calderon-Margalit 2009	1.7%	3.87 [1.53, 9.79]
Christensen 2015	1.9%	1.47 [0.62, 3.49]
Freeman 2018	0.4%	4.91 [0.69, 34.94]
Kallen 2013	96.1%	2.18 [1.93, 2.46]
Total (95% CI)	100.0%	2.19 [1.94, 2.47]

Heterogeneity: Tau^2 = 0.00; Chi^2 = 2.92, df = 3 (P = 0.40); I^2 = 0%
Test for overall effect: Z = 12.88 (P < 0. 00001)

E

Study or Subgroup	Weight	Odds Ratio IV, Random, 95% CI
Calderon-Margalit 2009	27.9%	4.33 [2.45, 7.65]
Freeman 2018	26.7%	2.02 [1.11, 3.68]
Laegreid 1992	8.7%	4.37 [1.07, 17.94]
Sutter-Dallay 2015	36.7%	1.89 [1.30, 2.75]
Total (95% CI)	100.0%	2.19 [1.94, 2.47]

Heterogeneity: Tau^2 = 0.11; Chi^2 = 6.70, df = 3 (P = 0.08); I^2 = 55%
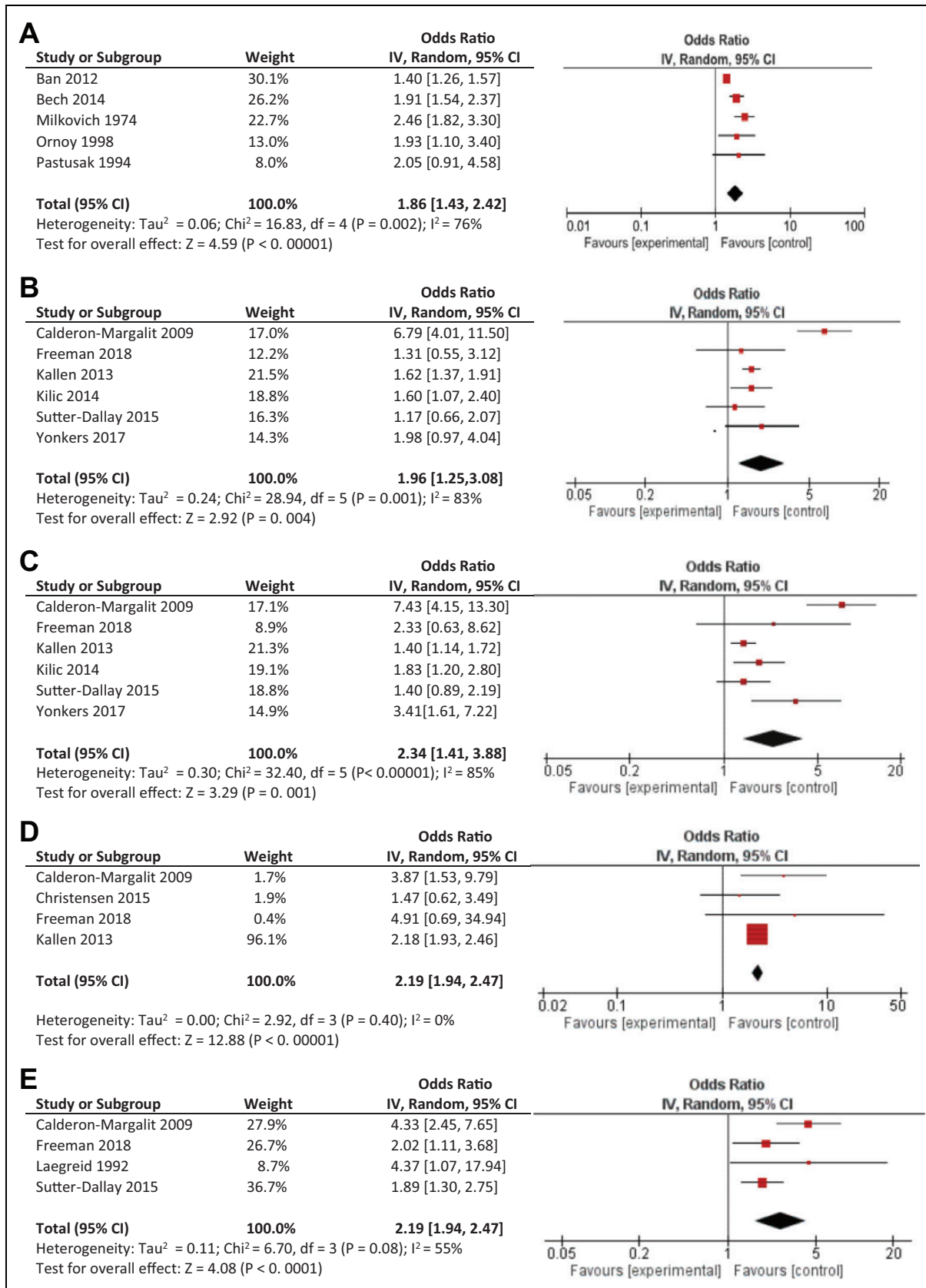Test for overall effect: Z = 4.08 (P < 0. 0001)

Figure 2. (A) (B), (C), (D) and (E). Significant adverse delivery outcomes following antenatal exposure to benzodiazepines. Pooled odds ratio for (A) spontaneous abortion, (B) preterm birth, (C) low birth weight, (D) low Apgar scores at 5 min, and (E) Neonatal Intensive Care Unit admission.

Table 1. Potential Moderators of the Effect of Antenatal Benzodiazepine Exposure on Delivery Outcomes.

Analysis	No. of Studies	Within Group — Odds Ratio or Mean Difference (95% CI)[a]	P value	$Q_{(df)}$ Within	Heterogeneity P Value	Heterogeneity I^2 (Percentage of Variance Explained)	$Q_{(df)}$ Between	Effect of Moderator P Value	Effect of Moderator I^2 (Percentage of Variance Explained)
Spontaneous abortion									
All studies	5	1.86 (1.43 to 2.42)	<0.001	16.83_4	0.002	76.0			
Study Quality									
Above quality threshold	4	1.85 (1.3 to 2.49)	<0.001	16.33_3	0.001	82.0	0.02_1	0.89	0.0
Below quality threshold	1	1.93 (1.10 to 3.40)	0.02						
Timing of BZD exposure									
Any time in pregnancy	3	2.30 (1.79 to 2.96)	<0.001	0.63_2	0.73	0.0	16.19_2	<0.001	87.6
First trimester	1	1.40 (1.26 to 1.57)	<0.001						
Up to 22 weeks gestation	1	1.91 (1.54 to 2.37)	<0.001						
Any adjusted data									
Adjusted findings	2	1.61 (1.19 to 2.18)	0.002	6.24_1	0.01	84.0	3.17_1	0.07	68.5
Unadjusted findings	3	2.30 (1.79 to 2.96)	<0.001	0.63_2	0.73	0.0			
Exposure to other psychotropics									
No, exclusion criteria	1	1.40 (1.26 to 1.57)	<0.001				14.96_1	<0.001	93.3
Not specified/assessed	4	2.07 (1.76 to 2.44)	<0.001	1.86_3	0.60	0.0			
Controlled for psychiatric diagnoses									
Yes, adjusted	1	1.91 (1.54 to 2.37)	<0.001				16.19_2	<0.001	87.6
Exclusion criteria	1	1.40 (1.26 to 1.57)	<0.001						
Not specified/assessed	3	2.30 (1.79 to 2.96)	<0.001	0.63_2	0.73	0.0			
Country									
North America	2	2.40 (1.82 to 3.17)	<0.001	0.17_1	0.68	0.0	3.63_2	0.16	44.9
Europe	2	1.61 (1.19 to 2.18)	0.002	6.24_1	0.01	84.0			
Other	1	1.93 (1.10 to 3.40)	0.02						
Preterm birth									
All studies	6	1.96 (1.25 to 3.08)	0.004	28.94_5	<0.001	83.0			
Study quality									
Above quality threshold	6	1.96 (1.25 to 3.08)	0.004	28.94_5	<0.001	83.0	0.45_1	0.50	0.0
Below quality threshold	0								
Timing of BZD exposure									
Any time in pregnancy	5	2.05 (1.05 to 3.99)	0.04	25.96_4	<0.001	85.0			
Second and/or third trimester	1	1.62 (1.37 to 1.91)	<0.001						
Any adjusted data									
Adjusted findings	6	1.96 (1.25 to 3.08)	0.004	28.94_5	<0.001	83.0			
Unadjusted findings	0								
Exposure to other psychotropics									
Yes, adjusted	3	1.41 (0.95 to 2.10)	0.09	1.31_2	0.52	0.0	1.81_1	0.18	44.9
Not specified/assessed	3	2.51 (1.20 to 5.27)	0.01	26.31_2	<0.001	92.0			
Controlled for psychiatric diagnoses									
Yes, adjusted	3	1.41 (0.95 to 2.10)	0.09	1.31_2	0.52	0.0	1.81_1	0.18	44.9
Not specified/assessed	3	2.51 (1.20 to 5.27)	0.01	26.31_2	<0.001	92.0			
Country									
North America	3	2.71 (0.96 to 7.65)	0.06	13.45_2	0.001	85.0	1.02_1	0.31	1.5

(continued)

Table 1. (continued)

Analysis	No. of Studies	Within Group — Odds Ratio or Mean Difference (95% CI)[a]	P value	$Q_{(df)}$ Within	Heterogeneity — P Value	I^2 (Percentage of Variance Explained)	Effect of Moderator — $Q_{(df)}$ Between	P Value	I^2 (Percentage of Variance Explained)
Europe	3	1.58 (1.36 to 1.83)	<0.001	1.14_2	0.57	0.0			
Other	0								
Birth weight									
All studies	5	−151.35 (−329.73 to 27.03)	0.10	31.90_4	<0.001	87.0			
Study quality									
Above quality threshold	4	−204.65 (−444.70 to 35.39)	0.09	23.02_4	<0.001	87.0	2.84_1	0.09	64.7
Below quality threshold	1	15.40 (−73.93 to 104.73)	0.74						
Timing of BZD exposure									
Any time in pregnancy	5	−151.35 (−329.73 to 27.03)	0.10	31.90_4	<0.001	87.0			
First trimester	0								
Second and/or third trimester	0								
Any adjusted data									
Adjusted findings	1	−137.77 (−226.52 to −49.02)	0.002				0.03_1	0.86	0.0
Unadjusted findings	4	−163.89 (−437.83 to 110.04)	0.24	30.10_3	<0.001	90.0			
Exposure to other psychotropics									
No, exclusion criteria	3	−143.47 (−478.13 to 191.20)	0.40	28.38_2	<0.001	93.0	0.00_1	0.98	0.0
Not specified/assessed	2	−147.14 (−231.77 to −62.51)	<0.001	0.47_1	0.49	0.0			
Controlled for psychiatric diagnoses									
Yes (exclusion criteria)	2	−147.14 (−231.77 to −62.51)	<0.001	0.47_1	0.49	0.0	0.00_1	0.98	0.0
No (not specified/assessed)	3	−143.47 (−478.13 to 191.20)	0.40	28.38_2	<0.001	93.0			
Country									
North America	2	−233.02 (−873.62 to 407.58)	0.48	22.29_1	<0.001	96.0	6.95_2	0.03	71.2
Europe	2	−147.14 (−231.77 to −62.51)	<0.001	0.47_1	0.49	0.0			
Other	1	15.40 (−73.93 to 104.73)	0.74						
Low birth weight									
All studies	6	2.34 (1.41 to 3.88)	0.001	32.40_4	<0.001	85.0			
Study quality									
Above quality threshold	6	2.34 (1.41 to 3.88)	0.001	32.40_4	<0.001	85.0			
Below quality threshold	0								
Timing of BZD exposure									
Any time in pregnancy	5	2.70 (1.42 to 5.15)	0.003	22.41_4	<0.001	82.0	3.62_1	0.06	72.4
Second and/or third trimester	1	1.40 (1.14 to 1.72)	0.001						
Any adjusted data									
Adjusted findings	6	2.34 (1.41 to 3.88)	0.001	32.40_4	<0.001	85.0			
Unadjusted findings	0								
Exposure to other psychotropics									
Yes, adjusted	3	2.06 (1.10 to 3.89)	0.02	4.16_2	0.13	52.0	0.17_1	0.68	0.0
Not specified/assessed	3	2.57 (1.11 to 5.95)	0.03	28.18_2	<0.001	93.0			
Controlled for psychiatric diagnoses									
Yes	3	2.06 (1.10 to 3.89)	0.02	4.16_2	0.13	52.0	0.17_1	0.68	0.0
No	3	2.57 (1.11 to 5.95)	0.03	28.18_2	<0.001	93.0			

(continued)

Table 1. (continued)

Analysis	No. of Studies	Within Group Odds Ratio or Mean Difference (95% CI)[a]	P value	$Q_{(df)}$ Within	Heterogeneity P Value	I^2 (Percentage of Variance Explained)	Effect of Moderator $Q_{(df)}$ Between	P Value	I^2 (Percentage of Variance Explained)
Country							9.89_1	0.002	89.9
North America	3	4.48 (2.28 to 8.82)	<0.001	4.08_2	0.13	51.0			
Europe	3	1.46 (1.23 to 1.74)	<0.001	1.30_2	0.52	0.0			
Other	0								
Gestational age									
All studies	5	−0.49 (−1.18 to 0.19)	0.16	44.35_4	<0.001	91.0			
Study quality							0.74_1	0.39	0.0
Above quality threshold	4	−0.62 (−1.76 to 0.52)	0.29	41.80_3	<0.001	93.0			
Below quality threshold	1	−0.10 (−0.40 to 0.20)	0.51						
Timing of BZD exposure									
Any time in pregnancy	5	−0.49 (−1.18 to 0.19)	0.16	44.35_4	<0.001				
First trimester	0								
Second and/or third trimester	0								
Any adjusted data							1.52_1	0.22	34.2
Adjusted findings	2	−1.47 (−3.82 to 0.88)	0.22	34.02_1	<0.001	97.0			
Unadjusted findings	3	0.02 (−0.30 to 0.35)	0.90	2.32_2	0.31	14.0			
Exposure to other psychotropics							0.33_1	0.57	0.0
Yes, exclusion criteria	2	−0.28 (−0.52 to −0.05)	0.02	0.37_1	0.54	0.0			
Not specified/assessed	3	−0.75 (−2.32 to 0.83)	0.35	43.95_2	<0.001	95.0			
Controlled for psychiatric diagnoses							0.33_1	0.57	0.0
Yes, exclusion criteria	2	−0.28 (−0.52 to −0.05)	0.02	0.37_1	0.54	0.0			
Not specified/assessed	3	−0.75 (−2.32 to 0.83)	0.35	43.95_2	<0.001	95.0			
Country							1.19_2	0.55	0.0
North America	2	−1.10 (−4.23 to 2.04)	0.49	35.73_1	<0.001	97.0			
Europe	2	−0.28 (−0.52 to −0.05)	0.02	0.37_1	0.54	0.0			
Other	1	−0.10 (−0.40 to 0.20)	0.51						

Note. GA = gestational age; SGA = small for gestational age.
[a] Pooled effect size estimated using random-effects model.
[b] Other countries include Australia, Israel, Singapore, and Taiwan.

829

be positive for slight bias after visual inspection of the funnel plot. Interestingly, using trim and fill, the adjusted *OR* was of the same magnitude but reached significance (*OR* = 1.38, 95% CI, 1.04 to 1.85, *P* = 0.03). The *induced abortion* outcome had too few studies for generation of a funnel plot.

Discussion

The aim of this work was to synthesize the literature and quantify potential risk of antenatal BZD exposure. Outcome data were found for 9 outcomes. Adverse effects following maternal use of BZDs during pregnancy were significantly associated with 6 outcomes overall, but 8 with an adjusted effect accounting for publication bias. There was significant heterogeneity for most outcomes, which suggests that confounding cannot be ruled out, although a consistent factor did not seem to emerge in the moderator analyses. Regardless, the magnitude of the majority of the pooled *OR*s approximated 2, which many experts believe is the threshold for clinical significance.[33] While the NICU outcome appears to be of clinical significance, which makes theoretical sense, it is reassuring that most of the effects do not appear to be alarming. Our work suggests much more investigation is needed into the potential adverse effects on maternal and neonatal outcomes following antenatal BZD exposure.

Strengths and Limitations

The study had several strengths, the most important of which was the rigorous methodology. Our search was comprehensive, and a broad range of perinatal outcomes was examined. A priori, we set out to explore potential sources of heterogeneity, regardless of significance. Interestingly, although there was significant heterogeneity, our a priori determined potential factors were not consistently significant but do suggest that we need to keep looking for other factors. We assessed study quality to ensure we based our conclusions on the best evidence. The fact that study quality was not a significant source of heterogeneity increases our confidence in the results with the majority of studies being above our quality threshold. Moreover, the findings are consistent with the antidepressant literature, but the magnitude of the *OR*s generally was higher, suggesting potentially stronger effects.

Study limitations stem primarily from methodological limitations of the included studies and represent the early state of scientific knowledge in the BZD literature and not the systematic review methods themselves. Specifically, all included studies were observational studies. Of these, 4/14 studies did not describe the indication for BZD use, 2/14 were based on women with epilepsy, and in 1 study, BZDs were used for obstetrical reasons. The studies that did not report indication for use of the BZD were typically the ones with the larger sample sizes. Further, the studies where indication was noted for a psychiatric disorder were based on women with mixed diagnoses, predominantly depression and anxiety.[10,26,27,29,31,32] While we did not find a significant effect for the moderator variable of diagnoses (or anxiety indication, data not shown), these factors indicate the study populations were likely heterogeneous. Psychiatric or other medical disorders (i.e., epilepsy) which may have required BZD use may themselves be associated with poor birth outcomes (e.g., preterm birth, delivery of a low birth weight infant).[13,34,35] A good example is the spontaneous abortion outcome. The subgroups with the "unspecified" groupings pooled to the highest *OR*s among the significant moderator sub-analyses. This is evidence for sample type explaining the between-study heterogeneity. Further, the Ban et al.[10] study, which controlled for illness, found that the women with illness who stopped using BZDs had a lower risk compared to those who continued, but acknowledged that the severity of the illness (which was not measured) may have affected the results. That is, women with more severe illness may have stayed on the drug, and thus, the illness, and not necessarily the drug, may have affected the risk. The above limitations are a result of confounding, that is, the BZD association is distorted because other factors that the study groups differ on, such as diagnosis or indication, can also be associated with adverse outcomes. Effects of confounding by indication can explain a portion or all of the associations found. Moreover, residual confounding can still affect results even after controlling for certain variables. For example, the Bech et al.[9] study statistically adjusted for a number of factors, but the authors acknowledged that the increased risk can still be attributed to "unmeasured confounding." The Calderon-Margalit et al.[23] study acknowledged that both the indication (i.e., anxiety), which was not measured, and residual confounding by unmeasured variables (such as illicit drug use) may have influenced the results. At least 6/14 studies directly discuss confounding by indication and 10/14 residual confounding as limitations to their results.

Another limitation in pharmacoepidemiology studies includes noncompliance with the treatment. Recording in a database when a study drug is prescribed does not necessarily mean that the drug was taken as prescribed. Kallen et al.[36] reviewed this literature and concluded that data based on prescriptions do not necessarily reflect accurate use of the drugs. The studies that use prescription databases are the largest ones in terms of sample size and thus typically dominate the meta-analyses. This can be seen in the low Apgar score at 5-minute analysis, where the Kallen et al.[5] study has a weight of 96.1% in the pooled *OR*, thereby contributing more to the weighted average than the other studies.[37] Moreover, in these large database studies, the time period of exposure may also not be accurate.[38]

Given the limited data, we grouped the BZDs without examining individual drugs; because they differ at least in half-life, this may also limit our conclusions. Moreover, the pattern of use was not known. Chronic versus intermittent use of BZDs, especially at higher doses and with shorter half-life drugs, is known to confer different consequences in adults such as higher likelihood of dependence.[39−41] BZD

dependence itself is linked to social determinants of health; these in turn may affect outcomes and thus confound results. In 6/14 studies (see Supplemental Table 1), the BZDs appear to have been prescribed or found in the medical record at least once, but it is unclear how often the drugs were used (e.g., other than during the first trimester). Furthermore, there is evidence that chronic versus occasional use has differing reliability implications, that is, self-report may be less reliable with occasional use.[42] Bech et al.[9] and Kilic et al.[11] reported an analysis of high versus low dose of all the drugs they studied (antiepileptics, including clonazepam); there were significant effects regardless of dose, although the magnitude of effects differed. Czeizel et al.[25] examined a wide dose range of diazepam from minimal dose up to 460 mg, while Laegreid et al.[26] described regular use of BZDs studied. In addition, known confounds for obstetrical outcomes were infrequently addressed. For example, maternal smoking[43] was not controlled in studies,[9,25,27,28] even if reported as significantly different between groups.[26]

We have previously reported that antidepressants are associated with adverse perinatal outcomes.[44] Although not often investigated, some reports found elevated risks of neonatal outcomes associated with concurrent use of sedatives/hypnotics and antidepressants compared to the risks associated with the individual medication class.[5,45] Whether this increased risk is related to the combination of medications or, perhaps, to more severe psychopathology (necessitating combinations of medications) is unknown. The studies in our meta-analysis were also heterogeneous regarding concurrent use of other psychotropic medications (in addition to BZDs) in pregnancy, including different medications with unknown doses and windows of exposure.

Finally, the Freeman et al.[31] study (a national registry of women housed in Massachusetts) and the Yonkers et al.[32] study (which recruited some women from Massachusetts) may have some overlap. There was an 8-month overlap during their subject recruitment period, and it is possible that some women may be registered in both studies. While this could potentially inflate the results, the results did not change when these 2 studies were excluded.

Conclusions

Anxiety disorders occur in up to 15% of pregnant women and are as common during pregnancy as depression; some experts argue more so.[46-49] Anxiety disorders are also associated with poor birth outcomes such as preterm birth and delivering a low birth weight infant.[13,34,35] Recommended treatments for anxiety disorders include both psychological and pharmacological interventions, but women are often reluctant to use medications during pregnancy for fear of negative effects on the developing fetus.[50] Our review found evidence for adverse effects of antenatal BZD use/exposure, but the actual clinical risks to fetal/neonatal health are not completely understood because statistical significance does not necessarily confer clinical significance. An adverse

outcome that is clinically relevant is NICU admission with the highest *OR* at 2.61. Although there were only 4 studies for this outcome, it is well-established that there is a risk of neonatal withdrawal effects following BZD exposure.[51] Chronic use of BZDs and use proximal to delivery is associated with neonatal toxicity/withdrawal, and the four studies in question reported the exposure to be anytime or throughout pregnancy. This is also consistent with the finding that Apgar at 5 min was lower in the exposed group, as those infants would more likely be admitted to the NICU. Other potential adverse outcomes such as spontaneous abortion are more difficult to address. Although the *OR* of 1.86 was statistically significant, many pregnancies result in spontaneous abortion and the clinical significance is unclear. The true rate of spontaneous abortion is unknown, but rates up to 20% and higher for older women have been reported.[52-54] Most important, however, is that women who are prescribed BZDs are also exposed to a number of possible confounding factors, which makes it premature to draw conclusions regarding causality. Pregnant women who take BZDs need to be followed closely and ideally by a perinatal expert who is familiar with the multiple potential confounds and consequences.

Most of the studies included in this meta-analysis examined risks associated with use of any BZD, with 2 studies focused on clonazepam, 1 on diazepam, and 1 on chlordiazepoxide. Given that duration of action for these BZDs varies, future research should examine whether there are differential risks for the various medications. It is imperative that future research use consistent methods. Potential confounds such as psychiatric diagnosis, type of medication, dose, and pattern of use should be controlled for in the analysis.[38,55]

Psychiatric illnesses remain stigmatized, yet are among the most prevalent illnesses in pregnancy. However, they are also undertreated, which can result in serious consequences for the mother and her family, including death by suicide.[56] The provision of effective treatment options is a public health priority. Given the potential clinical significance of the results, our meta-analysis suggests that pregnant women using BZDs should be treated as a high-risk obstetrical group.

ORCID iD

Sophie Grigoriadis, MD, MA, PhD, FRCPC 🄳 https://orcid.org/0000-0003-3461-6850

Supplemental Material

The Supplemental Material for this article is available online.

References

1. Leong C, Raymond C, Chateau D, et al. Psychotropic drug use before, during, and after pregnancy: a population-based study in a Canadian cohort (2001-2013). Can J Psychiatry. 2017; 62(8):543-550.

2. Petersen I, McCrea RL, Sammon CJ, et al. Risks and benefits of psychotropic medication in pregnancy: cohort studies based on UK electronic primary care health records. Health Technol Assess. 2016;20(23):1-176.

3. Epstein RA, Bobo WV, Shelton RC, et al. Increasing use of atypical antipsychotics and anticonvulsants during pregnancy. Pharmacoepidemiol Drug Saf. 2013;22(7):794-801.

4. Bobo WV, Davis RL, Toh S, et al. Trends in the use of antiepileptic drugs among pregnant women in the US, 2001-2007: a medication exposure in pregnancy risk evaluation program study. Paediatr Perinat Epidemiol. 2012;26(6):578-588.

5. Kallen B, Borg N, Reis M. The use of central nervous system active drugs during pregnancy. Pharmaceuticals. 2013;6(10): 1221-1286.

6. Bellantuono C, Tofani S, Di Sciascio G, Santone G. Benzodiazepine exposure in pregnancy and risk of major malformations: a critical overview. Gen Hosp Psychiatry. 2013;35(1):3-8.

7. Enato E, Moretti M, Koren G. The fetal safety of benzodiazepines: an updated meta-analysis. J Obstet Gynaecol Can. 2011; 33(1):46-48.

8. Dolovich LR, Addis A, Vaillancourt JM, Power JD, Koren G, Einarson TR. Benzodiazepine use in pregnancy and major malformations or oral cleft: meta-analysis of cohort and case-control studies. BMJ. 1998;317(7162):839-843.

9. Bech BH, Kjaersgaard MI, Pedersen HS, et al. Use of antiepileptic drugs during pregnancy and risk of spontaneous abortion and stillbirth: population based cohort study. BMJ. 2014;349: g5159.

10. Ban L, Tata LJ, West J, Fiaschi L, Gibson JE. Live and nonlive pregnancy outcomes among women with depression and anxiety: a population-based study. PLoS One. 2012;7(8): e43462.

11. Kilic D, Pedersen H, Kjaersgaard MI, et al. Birth outcomes after prenatal exposure to antiepileptic drugs—a population-based study. Epilepsia. 2014;55(11):1714-1721.

12. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283(15):2008-2012.

13. Grigoriadis S, Graves L, Peer M, et al. Maternal anxiety during pregnancy and the association with adverse perinatal outcomes: systematic review and meta-analysis. J Clin Psychiatry. 2018;79(5):17r12011.

14. McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). Evid Bas Library Inform Pract. 2016;5(1): 149-154.

15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

16. Ross LE, Grigoriadis S, Mamisashvili L, et al. Quality assessment of observational studies in psychiatry: an example from perinatal psychiatric research. Int J Methods Psychiatr Res. 2011;20(4):224-234.

17. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377-384.

18. Wells A, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis [accessed 2016 Jan 2.]. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

19. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926.

20. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-1457.

21. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-1558.

22. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-463.

23. Calderon-Margalit R, Qiu C, Ornoy A, Siscovick DS, Williams MA. Risk of preterm delivery and other adverse perinatal outcomes in relation to maternal use of psychotropic medications during pregnancy. Am J Obstet Gynecol. 2009; 201(6):579. e1-e8.

24. Christensen J, Pedersen HS, Kjaersgaard MI, et al. Apgar-score in children prenatally exposed to antiepileptic drugs: a population-based cohort study. BMJ Open. 2015;5(9): e007425.

25. Czeizel AE, Szegal BA, Joffe JM, Racz J. The effect of diazepam and promethazine treatment during pregnancy on the somatic development of human offspring. Neurotoxicol Teratol. 1999;21(2):157-167.

26. Laegreid L, Hagberg G, Lundberg A. The effect of benzodiazepines on the fetus and the newborn. Neuropediatrics. 1992; 23(1):18-23.

27. Milkovich L, van den Berg BJ. Effects of prenatal meprobamate and chlordiazepoxide hydrochloride on human embryonic and fetal development. N Engl J Med. 1974;291(24): 1268-1271.

28. Ornoy A, Arnon J, Shechtman S, Moerman L, Lukashova I. Is benzodiazepine use during pregnancy really teratogenic? Reprod Toxicol. 1998;12(5):511-515.

29. Pastuszak AL, Milich V, Can S, Chu I, Koren G. Prospective assessment of pregnancy outcome following first trimester exposure to benzodiazepines. Can J Clin Pharmacol. 1996; 3(4):167-171.

30. Sutter-Dallay AL, Bales M, Pambrun E, Glangeaud-Freudenthal NM, Wisner KL, Verdoux H. Impact of prenatal exposure to psychotropic drugs on neonatal outcome in infants of mothers with serious psychiatric illnesses. J Clin Psychiatry. 2015;76(7):967-973.

31. Freeman MP, Goez-Mogollon L, McInerney KA, et al. Obstetrical and neonatal outcomes after benzodiazepine exposure during pregnancy: results from a prospective registry of women with psychiatric disorders. Gen Hosp Psychiat. 2018; 53:73-79.

32. Yonkers KA, Gilstad-Hayden K, Forray A, Lipkind HS. Association of panic disorder, generalized anxiety disorder, and benzodiazepine treatment during pregnancy with risk of adverse birth outcomes. JAMA Psychiatry. 2017;74(11):1145-1152.

33. Einarson A. The importance of critical evaluation of the literature regarding safety of antidepressant use in pregnancy. Acta Psychiatr Scand. 2013;127(2):115-116.

34. Rose MS, Pana G, Premji S. Prenatal maternal anxiety as a risk factor for preterm birth and the effects of heterogeneity on this relationship: a systematic review and meta-analysis. Biomed Res Int. 2016;2016:8312158.

35. Ding XX, Wu YL, Xu SJ, et al. Maternal anxiety during pregnancy and adverse birth outcomes: a systematic review and meta-analysis of prospective cohort studies. J Affect Disord. 2014;159:103-110.

36. Kallen B, Nilsson E, Olausson PO. Antidepressant use during pregnancy: comparison of data obtained from a prescription register and from antenatal care records. Eur J Clin Pharmacol. 2011;67(8):839-845.

37. Deeks JJHJ, Altman DG, on behalf of the Cochrane Statistical Methods Group. Chapter 10: Analysing data and undertaking meta-analyses; 2019 [accessed 2019 Sept 30]. http://www.train ing.cochrane.org/handbook2011.

38. Grzeskowiak LE, Gilbert AL, Morrison JL. Exposed or not exposed? Exploring exposure classification in studies using administrative data to investigate outcomes following medication use during pregnancy. Eur J Clin Pharmacol. 2012;68(5):459-467.

39. Baldwin DS, Aitchison K, Bateson A, et al. Benzodiazepines: risks and benefits. A reconsideration. J Psychopharmacol. 2013;27(11):967-971.

40. Zandstra SM, Van Rijswijk E, Rijnders CA, et al. Long-term benzodiazepine users in family practice: differences from short-term users in mental health, coping behaviour and psychological characteristics. Fam Pract. 2004;21(3):266-269.

41. Hallfors DD, Saxe L. The dependence potential of short half-life benzodiazepines: a meta-analysis. Am J Public Health. 1993;83(9):1300-1304.

42. Cohen JM, Wood ME, Hernandez-Diaz S, Nordeng H. Agreement between paternal self-reported medication use and records from a national prescription database. Pharmacoepidemiol Drug Saf. 2018;27(4):413-421.

43. Phillips JK, Skelly JM, King SE, Bernstein IM, Higgins ST. Associations of maternal obesity and smoking status with perinatal outcomes. J Matern Fetal Neonatal Med. 2018;31(12): 1620-1626.

44. Ross LE, Grigoriadis S, Mamisashvili L, et al. Selected pregnancy and delivery outcomes after exposure to antidepressant medication: a systematic review and meta-analysis. JAMA Psychiatry. 2013;70(4):436-443.

45. Oberlander TF, Misri S, Fitzgerald CE, Kostaras X, Rurak D, Riggs W. Pharmacologic factors associated with transient neonatal symptoms following prenatal psychotropic medication exposure. J Clin Psychiatry. 2004;65(2):230-237.

46. Fairbrother N, Janssen P, Antony MM, Tucker E, Young AH. Perinatal anxiety disorder prevalence and incidence. J Affect Disord. 2016;200:148-155.

47. Martini J, Petzoldt J, Einsle F, Beesdo-Baum K, Hofler M, Wittchen HU. Risk factors and course patterns of anxiety and depressive disorders during pregnancy and after delivery: a prospective-longitudinal study. J Affect Disord. 2015;175: 385-395.

48. Lee AM, Lam SK, Sze Mun Lau SM, Chong CS, Chui HW, Fong DY. Prevalence, course, and risk factors for antenatal anxiety and depression. Obstet Gynecol. 2007;110(5):1102-1112.

49. Dennis CL, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br J Psychiatry. 2017;210(5):315-323.

50. Battle CL, Salisbury AL, Schofield CA, Ortiz-Hernandez S. Perinatal antidepressant use: understanding women's preferences and concerns. J Psychiatr Pract. 2013;19(6):443-453.

51. Newport DJ, Fernandez S, Juric S, Stowe Z. Psychopharmacology during pregnancy and lactation. In: Schatzberg AF, Nemeroff CB, editors. The American Psychiatric Association Publishing Textbook of Psychopharmacology; 2009. doi: 10. 1176/appi.books.9781585623860.as64.

52. The Society of Obstetricians and Gynaecologists of Canada (SOGC). Miscarriage and stillbirth [accessed 2019 Oct 7]. https://www.pregnancyinfo.ca/your-pregnancy/special-consid eration/miscarriage/.

53. Buss L, Tolstrup J, Munk C, et al. Spontaneous abortion: a prospective cohort study of younger women from the general population in Denmark. Validation, occurrence and risk determinants. Acta Obstetricia et Gynecologica Scandinavica. 2006; 85(4):467-475.

54. Nybo Andersen AM, Wohlfahrt J, Christens P, Olsen J, Melbye M. Maternal age and fetal loss: population based register linkage study. BMJ. 2000;320(7251):1708-1712.

55. Grzeskowiak LE, Gilbert AL, Morrison JL. Investigating outcomes following the use of selective serotonin reuptake inhibitors for treating depression in pregnancy: a focus on methodological issues. Drug safety. 2011;34(11): 1027-1048.

56. Grigoriadis S, Wilton AS, Kurdyak PA, et al. Perinatal suicide in Ontario, Canada: a 15-year population-based study. CMAJ. 2017;189(34):E1085-E1092.