PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2015; **24**: 934–942
Published online 4 June 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3798

ORIGINAL REPORT

Methodological considerations in assessing the effectiveness of antidepressant medication continuation during pregnancy using administrative data†

Sonja A. Swanson[1]*, Sonia Hernandez-Diaz[1], Kristin Palmsten[2], Helen Mogun[3], Mark Olfson[4] and Krista F. Huybrechts[3]

[1]*Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA*
[2]*Department of Pediatrics, University of California, San Diego*
[3]*Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School*
[4]*New York State Psychiatric Institute, New York; the Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York*

ABSTRACT

Purpose The decision whether to continue antidepressant use for depression during pregnancy requires weighing maternal and child risks and benefits. Little is known about the effectiveness of antidepressant therapy during pregnancy. The goal of this study is to evaluate whether standard administrative claims data can be used to evaluate the effectiveness of antidepressants.

Methods Using prescription and healthcare visit Medicaid claims (2000–2007), we identified 28 493 women with a depression diagnosis and antidepressant fill in the 90 days before their last menstrual period. Antidepressant continuation was defined based on prescription fills during the first trimester. Depression hospitalizations and deliberate self-harm served as measures of the effectiveness of treatment continuation during pregnancy. Propensity score and instrumental variable analyses were used to attempt to account for confounding.

Results Relative to women who discontinued antidepressant therapy, women who continued were more likely to have a depression inpatient stay (odds ratio [OR] = 2.2, 95% confidence interval [95%CI]: 2.0–2.4) and deliberate self-harm code (OR = 1.4, 95%CI: 0.7–2.7). Accounting for measured covariates in the propensity score analysis, including age, race, comorbidities, comedications, features of the depression diagnosis, and antidepressant class, led to slightly attenuated estimates (OR = 2.0, 95%CI: 1.8–2.2; OR = 1.1, 95%CI: 0.5–2.4). Similar associations were estimated in subgroups with different levels of baseline depression severity. Proposed preference-time, calendar-time-based, and geography-based instruments were unlikely to meet the required conditions for a valid analysis.

Conclusions Our findings suggest that either antidepressant medications do not reduce the risk of depression relapse in pregnant women, or that administrative data alone could not be used to validly estimate the effectiveness of psychotropic medications during pregnancy. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS—antidepressant; pregnancy; effectiveness; administrative data; confounding; pharmacoepidemiolgy

Received 9 February 2015; Revised 16 April 2015; Accepted 20 April 2015

INTRODUCTION

Major depressive disorder and related mood and anxiety disorders are common in women of childbearing age, thus antidepressant (AD) medications are taken during nearly one in 10 pregnancies.[1–3] As with other medications used to treat chronic illnesses, the decision to continue versus discontinue AD therapy during pregnancy requires women and their physicians to weigh substantiated and theoretical risks and benefits. In the absence of randomized trials, we rely on observational studies to estimate the effects of AD therapy on maternal and child health outcomes to inform clinical risk/benefit assessments.

Observational studies based on administrative data have provided considerable insights into the *safety* of AD therapy during pregnancy, including estimates for the risk of unintended adverse outcomes like congenital malformations, preeclampsia, preterm birth, and

* Correspondence to: S. A. Swanson, Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Avenue, Kresge 9th Floor, Boston, MA 02115, USA. E-mail: sswanson@hsph.harvard.edu

† Prior presentation: This work was presented in part at the International Conference on Pharmacoepidemiology and Therapeutic Risk Management in 2013 in Montreal, Quebec.

postpartum hemorrhage.[4–8] However, less attention has been devoted to evaluating AD effectiveness during pregnancy with claims data. Studying the *effectiveness* of psychotropic medications in an observational study is especially difficult because of confounding by indication: patients with more severe depression may be more likely to continue AD therapy and less likely to remit. As such, comparing patients who continue versus discontinue AD therapy without appropriate adjustment for confounding is prone to biased estimation of the clinical effects of continuing AD therapy.[9] While safety studies are not immune to confounding, depression severity is perhaps less strongly associated with safety outcomes than effectiveness outcomes, and therefore the magnitude of potential bias is likely smaller in safety studies.

To the extent that we are able to account for confounding through analytic techniques that adjust for measured[10,11] and potentially unmeasured[12–14] confounders, we may address the policy-relevant and clinically relevant question of how effective AD continuation is during pregnancy. To date, no study has used administrative data to assess this question, although administrative data have been used to assess the effectiveness of other treatments.[15,16] The goal of this study was to explore the feasibility of using administrative data to assess the effect of AD continuation during pregnancy on maternal mental health.

METHODS

Cohort

Data come from the Medicaid Analytic eXtract (MAX) for 46 US states and Washington, DC, collected in 2000–2007. These linked datasets did not include data from Montana, Connecticut, Michigan, or Arizona because of difficulties linking mothers and infants, or incomplete or unavailable data. The MAX dataset contains individual-level demographic and Medicaid enrollment information, physician services and hospitalizations and their accompanying diagnoses or procedures, and filled outpatient medication prescriptions. Details of the cohort have been provided elsewhere.[17] In brief, we identified all pregnancies with delivery-related claims among women ages 12–55 years and linked these pregnancies to liveborn infants. Last menstrual period (LMP) was estimated using a validated algorithm based on the delivery data combined with diagnostic codes indicative of preterm delivery.[18] We further required women to be Medicaid eligible, without supplementary private insurance or restricted benefits and without managed care plans that had incomplete claims information, for the 3 months prior to LMP, throughout pregnancy, and 1 month post delivery.

For the current study, we further restricted the analytic sample to the 28 493 pregnancies of women who, in the 90 days prior to LMP, had a depression diagnosis (indicated by the International Classification of Diseases [ICD]-9 codes for 296.2, 296.3, 296.9, 300.4, 309.0, 309.1, and 311 in an inpatient or outpatient claim) and filled at least one script for an AD medication. We also stratified this cohort to assess those on the lowest and highest levels of measured surrogates for depression severity, with high-risk patients ($n = 2437$) defined by having one or more psychiatric inpatient stays or a code for deliberate self-harm in the 90 days prior to LMP and low-risk patients ($n = 7274$) defined by being prescribed a selective serotonin reuptake inhibitor (i.e., a proxy for less severe or complicated depression)[19] and the absence of deliberate self-harm, psychiatric comorbidities (inpatient or outpatient), and comedications as defined below. Service utilization, comorbidities, and prior deliberate self-harm have been previously shown to be important indicators of depression severity and mental wellness.[20] The decision to stratify on severity was based on clinical trial evidence of AD therapy efficacy which suggests that AD therapy is likely more effective among patients with severe depression while there is limited evidence of their effectiveness in patients with mild or moderate depression.[21] Further, because prior research provides evidence for AD effectiveness during pregnancy in a high-risk population,[22] this stratification facilitates comparison of results with previous work.

Exposure

We assessed whether women continued their AD therapy using information on prescription fills during the first trimester. For our primary analyses, uninterrupted AD treatment throughout the first trimester allowing for a 14-day grace period defined AD continuation. In sensitivity analyses, we considered alternative definitions, including ones with shorter (7-day) and longer (30-day) grace periods and based on number of fills ($\geq$1 vs. 0 and $\geq$2 vs. 0–1).

Outcomes

Proxy outcomes of mental health status were assessed during pregnancy following our exposure assessment 91–280 days after LMP. We considered the number of inpatient depression stays, as indicated by the number of unique dates with inpatient depression diagnoses (see ICD-9 codes described earlier) in the claims records during this window. Because multiple stays were rare, we dichotomized this outcome as any versus none and two or more versus zero or one. We also considered

DOI: 10.1002/pds

deliberate self-harm, as indicated by ICD-9 codes E950.x-E958.x. Finally, we explored suicide ideation as an outcome, as a new ICD-9 code (V62.84) indicating ideation became available in 2006. However, the number of women who had this code was small ($n = 109$), and the percentage with the code restricted to women with an LMP in 2006–2007 was lower than expected (0.4%), and therefore did not provide a meaningful outcome.

Covariates

Several covariates measured in the 90 days prior to LMP were considered as potential confounders, including age, race/ethnicity, comorbidities, comedications, features of the AD treatment, and further proxies for depression severity. Comorbidities included psychiatric comorbidities associated with depression severity (anxiety disorder, bipolar disorder, personality disorder, alcohol use disorders, drug use disorders, and schizophrenia) and select physical comorbidities (cancer, diabetes, headache disorder, and epilepsy). Comedications included: anxiolytics, antipsychotics, lithium, benzodiazepines, barbiturates, opioids, anticonvulsants, and other hypnotics. We further considered the class of AD treatment (selective serotonin reuptake inhibitors, serotonin and norepinephrine reuptake inhibitors, tricyclics, bupropion, and other) , the specific depression ICD-9 code, and prior deliberate self-harm as important covariates. Because suicide ideation codes were only available for part of our study period, prior suicide ideation was not adjusted for in our primary analyses, but was used in sensitivity analyses to see whether this less sensitive but highly specific indicator may capture some of the residual confounding.[23]

Propensity score analyses

Propensity score (PS) methods were used to account for measured confounding.[10] These methods require creating a model predicting treatment based on pretreatment covariates described earlier. We then adjusted for the PS (i.e., the predicted probability of treatment continuation) in logistic regression models of the association between treatment continuation and maternal mental health outcomes. Primary analyses adjust for PS decile, but did not materially change when adjusting for more or fewer PS quantiles (e.g., quintiles, ventiles, or finer stratification) or continuous PS modeled quadratically. In addition to using predefined covariates, we also performed a high-dimensional propensity score (hdPS) analysis.[24] The hdPS algorithm evaluates thousands of diagnoses, procedures, and pharmacy claim codes to identify and prioritize covariates. We combined 200 empirically identified confounders with the investigator-identified covariates listed earlier.

Instrumental variable analyses

When valid, instrumental variable (IV) methods can account for both measured and unmeasured confounding.[12–14] These methods require a variable, known as an instrument, which meets three conditions: (1) the instrument is associated with AD therapy continuation; (2) the instrument does not cause the outcomes except through its relationship with AD therapy continuation; and (3) it is not confounded with the outcomes. To obtain a point estimate for the average treatment effect, a fourth effect homogeneity condition is also necessary.[12–14] Condition (1) can be assessed by estimating the association between the proposed instrument and treatment. It is critical that the association between the proposed instrument and exposure is sufficiently strong, as a weak association can ultimately amplify biases because of sample variability or violations of conditions (2)–(3).[25] Conditions (2)–(3) cannot be empirically verified so they require subject matter knowledge to justify their appropriateness. We used bias component plots[26] to assess whether imbalances in measured confounders would lead to more bias in an IV analysis versus non-IV analysis; see Appendix for details. Overall, these assessments aligned with previously suggested guidelines for conducting and reporting IV analyses.[27,28] We considered adaptations of three previously proposed instruments,[29,30] including ones based on prescriber preference for continuing AD therapy during pregnancy, calendar time around the three Food and Drug Administration warnings regarding AD use, and geographic variation in the proportion of women who continue AD treatment during pregnancy by state, as a previous study from this cohort indicated variability in treatment across these domains.[3] Details of the proposed instruments' definitions are provided in the Appendix.

RESULTS

Descriptive analyses

The distribution of pre-LMP covariates by treatment group is provided in Table 1. Women who continued AD therapy were more likely to have a comorbid psychiatric disorder and be concomitantly prescribed other psychotropic medications. This was observed overall and within both the high-risk and low-risk subgroups. The high-risk cohort generally had higher proportions of comorbidities and comedications relative to the full cohort.

DOI: 10.1002/pds

Table 1. Covariate distribution prior to the last menstrual period by antidepressant continuation for the full cohort and the high and low risk subcohorts, *n* (%) unless otherwise noted

	Full cohort		High risk subcohort		Low risk subcohort	
	Discontinued AD (*n* = 24 103)	Continued AD (*n* = 4390)	Discontinued AD (*n* = 2033)	Continued AD (*n* = 404)	Discontinued AD (*n* = 6449)	Continued AD (*n* = 825)
Demographics						
White	16 882 (70.0)	3515 (80.1)	1315 (64.7)	322 (79.7)	4481 (69.5)	666 (80.7)
Age (years)						
12–17	2205 (9.2)	274 (6.2)	294 (14.5)	39 (9.7)	753 (11.7)	84 (10.2)
18–24	10 931 (45.4)	1360 (31.0)	944 (46.4)	134 (33.2)	3140 (48.7)	300 (36.4)
25–34	9211 (38.2)	2146 (48.9)	683 (33.6)	182 (45.1)	2168 (33.6)	352 (42.7)
35+	1756 (7.3)	610 (13.9)	112 (5.5)	49 (12.1)	388 (6.0)	89 (10.8)
Psychiatric inpatient stays						
Mean (SD)	0.1 (0.5)	0.2 (0.6)	1.7 (1.0)	1.8 (1.1)	NA	NA
Any	1977 (8.2)	393 (9.0)	1977 (97.3)	393 (97.3)	NA	NA
Psychiatric severity						
Deliberate self-harm	163 (0.7)	33 (0.8)	163 (8.0)	33 (8.2)	NA	NA
Anxiety disorder	159 (0.7)	55 (1.3)	21 (1.0)	13 (3.2)	NA	NA
Bipolar disorder	1012 (4.2)	240 (5.5)	311 (15.3)	78 (19.3)	NA	NA
Personality disorder	458 (1.9)	149 (3.4)	256 (12.6)	73 (18.1)	NA	NA
Alcohol use disorder	2231 (9.3)	455 (10.4)	872 (42.9)	183 (45.3)	NA	NA
Drug use disorder	1493 (6.2)	329 (7.5)	594 (29.2)	133 (32.9)	NA	NA
Schizophrenia	215 (0.9)	53 (1.2)	91 (4.5)	25 (6.2)	NA	NA
Antidepressant class						
SSRI	18 164 (75.4)	3312 (75.4)	1563 (76.9)	317 (78.5)	6449 (100.0)	825 (100.0)
SNRI	2952 (12.2)	894 (20.4)	254 (12.5)	92 (22.8)	NA	NA
TCA	1275 (5.3)	353 (8.0)	103 (5.1)	28 (6.9)	NA	NA
Bupropion	3026 (12.6)	643 (14.6)	236 (11.6)	60 (14.9)	NA	NA
Other	3790 (15.7)	877 (20.0)	520 (25.6)	110 (27.2)	NA	NA
Specific depression diagnosis (ICD code)						
Depressive disorder not otherwise specified (311)	13 778 (57.2)	2357 (53.7)	1339 (65.9)	289 (71.5)	4152 (64.4)	514 (62.3)
Major depressive disorder single episode (296.2)	3597 (14.9)	661 (15.1)	683 (33.6)	107 (26.5)	740 (11.5)	119 (14.4)
Major depressive disorder recurrent episode (296.3)	6450 (26.8)	1468 (33.4)	866 (42.6)	184 (45.5)	1070 (16.6)	186 (22.5)
Other and unspecified episodic mood disorder (296.9)	80 (0.3)	11 (0.3)	13 (0.6)	NA	NA	NA
Dysthymia (300.4)	3360 (13.9)	637 (14.5)	302 (14.9)	78 (19.3)	857 (13.3)	85 (10.3)
Adjustment disorder with depressed mood (309.0)	1021 (4.2)	129 (2.9)	140 (6.9)	18 (4.5)	274 (4.2)	22 (2.7)
Prolonged depressive reaction (309.1)	106 (0.4)	12 (0.3)	NA	NA	35 (0.5)	NA
Comedications						
Anxiolytics	585 (2.4)	156 (3.6)	55 (2.7)	21 (5.2)	NA	NA
Antipsychotics	3031 (12.6)	753 (17.2)	545 (26.8)	162 (40.1)	NA	NA
Lithium	237 (1.0)	61 (1.4)	51 (2.5)	NA	NA	NA
Benzodiazepine	4357 (18.1)	1305 (29.7)	472 (23.2)	167 (41.3)	NA	NA
Barbiturates	482 (2.0)	113 (2.6)	36 (1.8)	15 (3.7)	NA	NA
Opioid	6226 (25.8)	1325 (30.2)	662 (32.6)	167 (41.3)	NA	NA
Anticonvulsants	2249 (9.3)	663 (15.1)	381 (18.7)	117 (29.0)	NA	NA
Other hypnotics	2505 (10.4)	592 (13.5)	293 (14.4)	76 (18.8)	NA	NA
Non-psychiatric comorbidities						
Cancer	582 (2.4)	133 (3.0)	70 (3.4)	19 (4.7)	103 (1.6)	17 (2.1)
Diabetes	412 (1.7)	106 (2.4)	66 (3.2)	22 (5.4)	78 (1.2)	14 (1.7)
Headache disorder	1124 (4.7)	258 (5.9)	140 (6.9)	40 (9.9)	134 (2.1)	23 (2.8)
Epilepsy	122 (0.5)	42 (1.0)	30 (1.5)	13 (3.2)	11 (0.2)	NA

AD, antidepressant, SD, standard deviation; SSRI, selective serotonin reuptake inhibitors; SNRI, serotonin and norepinephrine reuptake inhibitors; TCA, tricyclic antidepressants; ICD, International Classification of Diseases.
NA, cell counts <11 are suppressed based on the Centers for Medicare and Medicaid Services requirements.

DOI: 10.1002/pds

Later in pregnancy, the number of women with an inpatient stay with a depression diagnosis was 2077 (8.6%) and 743 (16.9%) for those who discontinued versus continued AD treatment, respectively; 237 (1.0%) and 91 (2.1%) had two or more inpatient stays. Among the high-risk subgroup, 305 (15.0%) and 106 (26.2%) of women who discontinued versus continued AD treatment had at least one inpatient stay with a depression diagnosis; 71 (3.5%) and 25 (6.2%) had two or more stays. Among the low-risk subgroup, 445 (6.9%) and 116 (14.1%) had an inpatient stay with depression; 29 (0.5%) and 12 (1.5%) had two or more. Deliberate self-harm was relatively rarely recorded, with only 50 (0.2%) women engaging in deliberate self-harm; because of its rarity, we do not present results for this outcome within subgroups (data use agreements do not permit us to display cell sizes <11).

Propensity score analyses

In Table 2, we provide increasing levels of covariate-adjusted estimates for the association between AD continuation and maternal mental health outcomes. Women who continued AD therapy (and who appeared more severely depressed *a priori*) were more likely to have a depression inpatient stay than women who discontinued AD therapy (odds ratio [OR] = 2.2, 95% confidence interval [95%CI]: 2.0–2.4). Adjustment for prespecified covariates only slightly attenuated the crude estimate (OR = 2.0, 95%CI: 1.8–2.2); further adjustment using the hdPS analysis decreased the OR estimate to 1.8 (95%CI 1.7–2.0). Similar patterns were seen for other mental health outcomes, although the low risk of deliberate self-harm led to wider confidence intervals. Strong associations that were only slightly attenuated when adjusting for covariates were also seen in both the high-risk and low-risk subgroups (Table 3), although the association appeared stronger in the low-risk subgroup. Prespecified covariates were reasonably predictive of outcomes (Supplemental Table 1). Defining exposure based on number of fills (1+: OR 2.2, 95% CI: 2.0–2.4; 2+: OR 2.2, 95%CI: 2.0–2.5) or as continuous use with shorter (7-day: OR 1.8, 95%CI: 1.6–2.0) or longer (30-day: OR 2.0, 95%CI: 1.8–2.2) grace periods resulted in similar estimates when adjusting for a PS based on prespecified covariates. Models fit with only women with LMP dates in 2006–2007 showed no further attenuation when adjusted for suicide ideation. Post-hoc adjustment for AD dose or by recorded severity of depression episodes (as indicated by the fifth digit of the ICD-9 code) did not meaningfully change results, although these measures were not consistently reported (data not shown).

Table 2. Propensity score adjusted odds ratios for antidepressant continuation predicting inpatient depression stays and deliberate self-harm with increasing levels of covariate adjustments

	1 or more inpatient depression days	2 or more inpatient depression days	Deliberate self-harm
Crude	2.2 (2.0, 2.4)	2.1 (1.7, 2.7)	1.4 (0.7, 2.7)
Model 1[a]	2.1 (1.9, 2.3)	2.1 (1.6, 2.7)	1.5 (0.7, 3.1)
Model 2[b]	2.1 (1.9, 2.3)	2.0 (1.6, 2.6)	1.3 (0.7, 2.7)
Model 3[c]	2.0 (1.8, 2.2)	1.8 (1.4, 2.3)	1.2 (0.6, 2.6)
Model 4[d]	2.0 (1.8, 2.2)	1.7 (1.3, 2.1)	1.1 (0.5, 2.4)
hdPS	1.8 (1.7, 2.0)	1.7 (1.3, 2.2)	NA

hdPS, high-dimensional propensity score.
[a]Adjusts for age and race.
[b]Adjusts for age, race, and psychiatric comorbidities.
[c]Adjusts for age, race, psychiatric comorbidities, class of antidepressant, and specific depression diagnosis.
[d]Adjusts for age, race, psychiatric comorbidities, class of antidepressant, specific depression diagnosis, other medications, other comorbidities, and number of outpatient claims.

Table 3. Propensity score adjusted odds ratios for antidepressant continuation predicting inpatient depression stays and deliberate self-harm with increasing levels of covariate adjustments in the high risk and low risk subcohorts

	High risk subcohort		Low risk subcohort	
	1 or more inpatient depression stays	2 or more inpatient depression stays	1 or more inpatient depression stays	2 or more inpatient depression stays
Crude	2.0 (1.6, 2.6)	1.8 (1.1, 2.9)	2.2 (1.8, 2.7)	3.3 (1.7, 6.4)
Model 1[a]	1.8 (1.4, 2.4)	1.6 (1.0, 2.7)	2.2 (1.8, 2.8)	3.2 (1.6, 6.4)
Model 2[b]	1.8 (1.4, 2.4)	1.6 (1.0, 2.7)	2.2 (1.8, 2.8)	3.2 (1.6, 6.4)
Model 3[c]	1.8 (1.3, 2.3)	1.4 (0.9, 2.3)	2.1 (1.7, 2.7)	3.0 (1.5, 6.0)
Model 4[d]	1.7 (1.3, 2.2)	1.2 (0.7, 2.0)	2.1 (1.7, 2.7)	2.9 (1.4, 5.8)

[a]Adjusts for age and race.
[b]Adjusts for age, race, and psychiatric comorbidities in the high risk subcohort. For low risk subcohort, this is the same as Model 1.
[c]Adjusts for age, race, psychiatric comorbidities, class of antidepressant, and specific depression diagnosis in the high risk subcohort. For low risk subcohort, adjusts for age, race, and specific depression diagnosis.
[d]Adjusts for age, race, psychiatric comorbidities, class of antidepressant, specific depression diagnosis, other medications, other comorbidities, and number of outpatient claims in the high risk subcohort. For the low risk subcohort, adjusts for age, race, specific depression diagnosis, other comorbidities, and number of outpatient claims.

Instrumental variable analyses

We did not identify any variable that satisfied the instrumental conditions. We first considered a provider's preference toward AD continuation during pregnancy, approximated as the provider's prescribing decision for their prior eligible patient. However, individual providers rarely saw enough eligible patients during our study period: thus, we could only define the proposed instrument for approximately one-third of patients. For women who continued AD therapy, the prescriber across AD refills often switched, while for women who discontinue AD therapy it is by definition unclear

DOI: 10.1002/pds

whether the provider who filled the previous prescription(s) aided in the decision to discontinue treatment (Supplemental Table 2). As such, the proposed instrument was only weakly associated with treatment (risk difference [RD] = 0.074, *F*-statistic = 40.2). The proposed calendar-time instruments were even weaker (2003 warning: RD = −0.004, *F*-statistic = 0.1; 2005 warning: RD = −0.009, *F*-statistic = 0.8; 2006 warning: RD = 0.001, *F*-statistic < 0.1). The proposed geography-based instrument was somewhat stronger (RD = 0.059, *F*-statistic = 150.9). Regardless, all proposed instruments failed to provide better balance of measured covariates and therefore possibly unmeasured covariates as well (Figure 1): IV analyses performed using any of these proposed instruments would be substantially more biased than non-IV analyses if we failed to account for these measured confounders or unmeasured confounders of similar strength. For these reasons, we did not perform an IV analysis to estimate the effect of AD continuation.

DISCUSSION

Pregnancy is a particularly difficult, but important, period of life to study the safety and effectiveness of medications.[31] To make an informed decision about whether to continue or discontinue AD therapy, pregnant women and their physicians should ideally have an evidence-based understanding of the risks and benefits. While several studies have illuminated the possible risks, we now add a third study to the inconsistent findings on potential benefits.[22,32] Our results failed to find a benefit. Because of the known risks of confounding by indication and limitations with available outcome proxies available in administrative data, we speculate that these results are explained by systematic biases. Alternatively, serotonin availability and transmission are affected by hormonal changes,[33] and dose adjustment of AD medication during pregnancy has proven challenging,[34] thus we should not definitively rule out the possibility that AD medications are rendered less effective during pregnancy. In the following, we discuss the methodological lessons from this study and outline considerations for future studies of this important public health question.

The two previous observational studies that have assessed the association between AD therapy during pregnancy and relapse yielded conflicting findings: Cohen and colleagues[22] found evidence suggesting discontinuing AD therapy led to relapse, while Yonkers and colleagues[32] found no benefits associated with AD use in pregnancy. These studies were conducted in highly selected populations that excluded women with evidence of substance use, bipolar disorder, and psychotic disorders, and recruited from specialty clinics with expertise in treating psychiatric illness during pregnancy, and obstetrical practices and hospital-based clinics in New England. Interestingly, the confounding structure we observed and were *a priori* concerned about in the current study was not present in either of

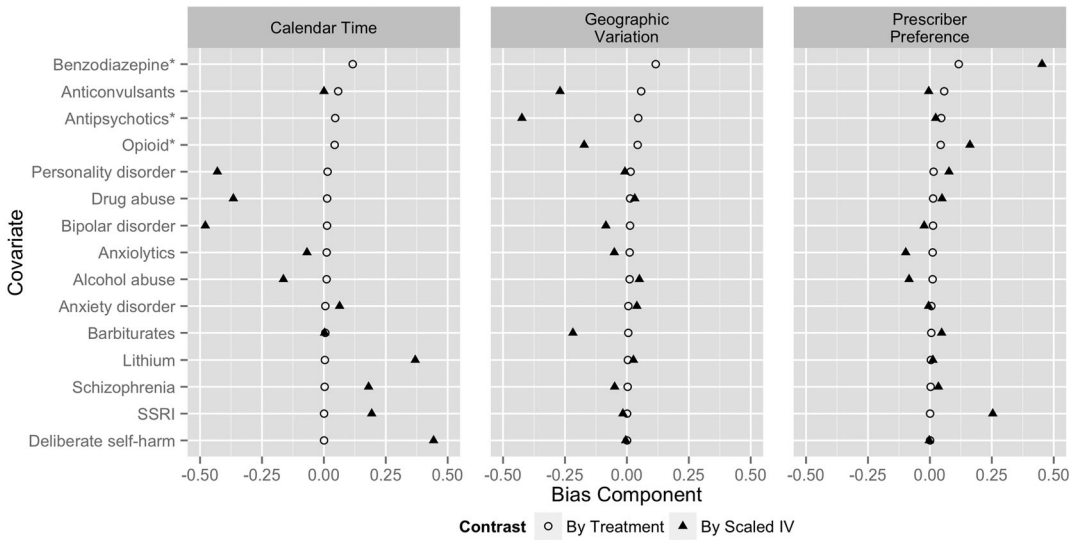

Figure 1. Bias component plots for three proposed instruments. Plots display the non-shared bias components in an IV analysis (i.e., by scaled IV) and non-IV analysis (i.e., by treatment) for select covariates. The *x*-axis is the prevalence difference in the covariate (scaled by the strength of the proposed instrument for the IV); greater imbalances indicate greater amounts of bias in an analysis that fails to account for the covariate. Asterisks indicate that the imbalances for benzodiazepine, antipsychotic, and opioid medications across levels of the calendar time proposed instrument, and benzodiazepines for the geographic variation proposed instrument, resulted in bias components larger than the scale of the graphs presented here. IV, instrumental variable; SSRI, selective serotonin reuptake inhibitor

DOI: 10.1002/pds

these studies: even in unadjusted results, both studies found that AD use during pregnancy appeared protective (with wide confidence intervals) against relapse defined by timed assessments with a structured diagnostic interview. Given that both studies occurred in clinical settings—one of which included sites specializing in treatment of psychiatric disorders during pregnancy—the decision to discontinue treatment in these settings likely reflects a mutual decision made under careful physician advisement. As our study relied on prescription fills and claims from a wide range of treatment settings, it is possible that decisions to discontinue AD therapy were more frequently made without the physician's knowledge or informed by signs not captured in claims data. Therefore, it is possible that the factors driving these decisions were less amenable to modeling in claims data.

When prescription decisions are based on quantifiable biological indicators, observational studies may be able to provide valid estimates for treatment effects after adjusting for measured confounders. For example, observational studies have been used to replicate results from randomized trials for the effectiveness of statin therapy and antiretroviral therapy, where prescription decisions are largely based on monitoring cholesterol levels and CD4 count, respectively.[15,16] For psychotropic medications, prescription decisions encompass subjective clinical assessments, thus some strong predictors of both the prescription decisions and mental health outcomes may be less amenable to investigation through claims or medical records that do not provide access to depression rating scales. Despite using multiple proxies for depression severity and other relevant confounders available in Medicaid claims, we presumably could not overcome confounding bias using techniques that require robust measurement of confounders. Moreover, our analyses do not address time-varying confounding. Time-varying confounding is especially problematic for studying discontinuation of psychotropic medications, as abrupt changes in medication are ill-advised and patients may reverse their decision to discontinue or adjust dosage when faced with withdrawal effects.[35] However, g-methods that could appropriately handle time-varying confounding require accurate measurement of the influences on treatment decisions repeatedly over the unit of time that treatment decisions occur, which is generally unattainable in studies based on monthly prescription fills when the relevant unit of time is on the order of days or weeks rather than months.

When faced with unmeasured confounding, IV methods offer an alternative possibility for estimating causal effects. Based on those previously used in other pharmacoepidemiology studies,[29,30] we proposed three possible instruments. While none of these met the instrumental conditions to our satisfaction, the reasons may be instructive. First, our difficulty in identifying a sufficiently strong preference-based or calendar-time-based instrument reminds us that women facing the decision to continue AD therapy are weighing many possible risks and benefits, and a single source of variation like a new Food and Drug Administration warning or the preference of the prescriber is unlikely to greatly influence such a decision. Moreover, approximating the prescriber's preference was especially challenging in the current study because (1) physicians rarely saw more than one Medicaid-financed pregnant patient facing the decision; and (2) we only had information on the prescriber associated with refills of a prescription, thus we may have been unable to adequately identify the physicians who were informing decisions to discontinue AD therapy. When we looked at state-level geographic variation, we were able to propose a sufficiently strong instrument, but this variable was even more confounded than our treatment. Specifically, the imbalance of other psychotropic medication use across patients who continued versus discontinued was smaller than the imbalance across patients living in states with high versus low continuation proportions. Together, these analyses serve as another reminder of the additional methodological challenges in using IV analyses to study medication use during pregnancy, and studying the effectiveness of psychotropic medications in claims data in general. Previously proposed instruments in pharmacoepidemiology may work better for studies where the influences on prescribing decisions are known to be limited (although perhaps unmeasured), or for studying medications used to treat conditions that would be overseen during pregnancy by a single specialist.

While our methodological focus has been on confounding, the proxies used to assess maternal mental health status have additional limitations that may explain our results. Although service utilization and self-harm may approximate severe mental health outcomes, they may fail to account for non-service-based changes in mental health status. Medically treated self-harm is an imperfect assessment of AD effectiveness because of the known adverse effects of AD therapy on suicide-related behaviors in youth and young adults.[36] This issue may be less of a concern in prevalent AD users, but it is unknown whether the hormonal changes associated with pregnancy create vulnerabilities to self-harm potentiating antidepressant effects. Nevertheless, greater contact with health care providers poses another potential source of confounding

DOI: 10.1002/pds

for both sets of outcomes. Women with continuous access to treatment may receive more consistent AD prescriptions and regular interactions with their providers that lower their clinical threshold for referral for an inpatient stay. Such explanations are supported by the fact that the two previous studies that directly assessed relapse did not have much apparent confounding by indication.

Methodologically, these results highlight the difficulties of assessing the effectiveness of psychiatric therapeutic strategies, and we have outlined some of the specific limitations of administrative data for addressing these important questions. Given the gravity of the broader question of whether women taking AD medications should continue use during pregnancy, and the limitations of currently available data in quantifying the possible maternal mental health benefits, a randomized trial would be highly informative. At minimum, the current study underscores the need for frequent and careful assessments of maternal mental health throughout pregnancy in order to understand the effectiveness of AD therapy using observational studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

KEY POINTS

- Observational studies based on administrative data have proved useful for assessing the safety of antidepressant medication use during pregnancy. However, estimating the effectiveness of antidepressant medication use in treating depression during pregnancy is challenging due to the lack of sufficiently sensitive measures of depression severity that serve as confounders and outcome proxies.
- While valid instrumental variable analyses may circumvent bias due to unmeasured confounding, the most commonly proposed types of instruments used by epidemiologists failed to meet the necessary conditions in this study.

ETHICS STATEMENT

The research was approved by the Institutional Review Board of Brigham and Women's Hospital. The Institutional Review Board granted a waver of informed consent.

ACKNOWLEDGEMENTS

This study was supported by the Agency for Healthcare Research and Quality (R01 HSO18533) and by the National Institutes of Health (R01 AI102634 and R01 MH100216). Krista Huybrechts is supported by a career development grant from the National Institute of Mental Health (K01 MH099141).

REFERENCES

1. Cooper WO, Willy ME, Pont SJ, Ray WA. Increasing use of antidepressants in pregnancy. *Am J Obstet Gynecol* Jun 2007; **196**(6):544 e541–e545.
2. Andrade SE, Raebel MA, Brown J, *et al.* Use of antidepressant medications during pregnancy: a multisite study. *Am J Obstet Gynecol* Feb 2008; **198**(2): 194 e191–e195.
3. Huybrechts KF, Palmsten K, Mogun H, *et al.* National trends in antidepressant medication treatment among publicly insured pregnant women. *Gen Hosp Psychiatry* May-Jun 2013; **35**(3): 265–271.
4. Wurst KE, Poole C, Ephross SA, Olshan AF. First trimester paroxetine use and the prevalence of congenital, specifically cardiac, defects: a meta-analysis of epidemiological studies. *Birth Defects Res A Clin Mol Teratol*. Mar 2010; **88**(3): 159–170.
5. Palmsten K, Setoguchi S, Margulis AV, Patrick AR, Hernandez-Diaz S. Elevated risk of preeclampsia in pregnant women with depression: depression or antidepressants?. *Am J Epidemiol* May 15 2012; **175**(10): 988–997.
6. Yonkers KA, Norwitz ER, Smith MV, *et al.* Depression and serotonin reuptake inhibitor treatment as risk factors for preterm birth. *Epidemiology* Sep 2012; **23**(5): 677–685.
7. Palmsten K, Hernandez-Diaz S, Huybrechts KF, *et al.* Use of antidepressants near delivery and risk of postpartum hemorrhage: cohort study of low income women in the United States. *BMJ* 2013; **347**: f4877.
8. Huybrechts KF, Palmsten K, Avorn J, *et al.* Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med* Jun 19 2014; **370**(25): 2397–2407.
9. Walker AM. Confounding by indication. *Epidemiology* Jul 1996; **7**(4): 335–336.
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**(1): 41–55.
11. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* (*Cambridge*, *Mass.*) 2009; **20**(4): 512–522.
12. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**(434): 444–455.
13. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream?. *Epidemiology* Jul 2006; **17**(4): 360–372.
14. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Community Stat* 1994; **23**: 2379–2412.
15. Danaei G, Rodriguez LA, Cantero OF, Logan R, Hernan MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res* Feb 2013; **22**(1): 70–96.
16. Collaboration H-C, Ray M, Logan R, *et al.* The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *Aids* Jan 2 2010; **24**(1): 123–137.
17. Palmsten K, Huybrechts KF, Mogun H, *et al.* Harnessing the medicaid analytic eXtract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS One* 2013; **8**(6): e67405.
18. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf* Jan 2013; **22**(1): 16–24.
19. Miller M, Pate V, Swanson SA, Azrael D, White A, Sturmer T. Antidepressant class, age, and the risk of deliberate self-harm: a propensity score matched cohort study of SSRI and SNRI users in the USA. *CNS Drugs* Jan 2014; **28**(1): 79–88.
20. Kessler RC, Berglund P, Demler O, *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* Jun 18 2003; **289**(23): 3095–3105.
21. Fournier JC, DeRubeis RJ, Hollon SD, *et al.* Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* Jan 6 2010; **303**(1): 47–53.
22. Cohen LS, Altshuler LL, Harlow BL, *et al.* Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. *JAMA*. Feb 1 2006; **295**(5): 499–507.
23. Miller M, Swanson SA, Azrael D, Pate V, Sturmer T. Antidepressant dose, age, and the risk of deliberate self-harm. *JAMA* Jun 2014; **174**(6): 899–909.

24. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* Jul 2009; **20**(4): 512–522.
25. Bound J, Jaeger D, Baker R. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 1995; **90**(430): 443–450.
26. Jackson J, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable analyses. *Epidemiology* (*Cambridge*, *Mass*.). in press.
27. Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* May 2013; **24**(3): 370–374.
28. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *The Int J biostatistics* 2007; **3**(1): Article 14.
29. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* Jun 2011; **64**(6): 687–700.
30. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* May 2013; **24**(3): 363–369.
31. Palmsten K, Hernandez-Diaz S. Can nonrandomized studies on the safety of antidepressants during pregnancy convincingly beat confounding, chance, and prior beliefs?. *Epidemiology* Sep 2012; **23**(5): 686–688.
32. Yonkers KA, Gotman N, Smith MV, *et al.* Does antidepressant use attenuate the risk of a major depressive episode in pregnancy?. *Epidemiology* Nov 2011; **22**(6): 848–854.
33. Lokuge S, Frey BN, Foster JA, Soares CN, Steiner M. Depression in women: windows of vulnerability and new insights into the link between estrogen and serotonin. *J Clin Psychiatry* Nov 2011; **72**(11): e1563–e1569.
34. Sit D, Perel JM, Luther JF, Wisniewski SR, Helsel JC, Wisner KL. Disposition of chiral and racemic fluoxetine and norfluoxetine across childbearing. *J Clin Psychopharmacol* Aug 2010; **30**(4): 381–386.
35. Schatzberg AF, Blier P, Delgado PL, Fava M, Haddad PM, Shelton RC. Antidepressant discontinuation syndrome: consensus panel recommendations for clinical management and additional research. *J Clin Psychiatry* 2006; **67**(Suppl 4): 27–30.
36. Stone M, Laughren T, Jones ML, *et al.* Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ* 2009; **339**: b2880.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

DOI: 10.1002/pds