Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2015 September ; 24(9): 934–942. doi:10.1002/pds.3798.

Methodological Considerations in Assessing the Effectiveness of Antidepressant Medication Continuation during Pregnancy using Administrative Data

Sonja A. Swanson, Sonia Hernandez-Diaz, Kristin Palmsten, Helen Mogun, Mark Olfson, and Krista Huybrechts

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 account for confounding.

Results—Relative to women who discontinued antidepressant therapy, women who continued were more likely to have a depression inpatient stay (OR=2.2, 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-, calendar-time-, 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 cannot be used to validly estimate the effectiveness of psychotropic medications during pregnancy.

Keywords

antidepressant; pregnancy; effectiveness; administrative data; confounding

Corresponding Author: Sonja A. Swanson, Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Avenue, Kresge 9th Floor, Boston, Massachusetts 02115, sswanson@hsph.harvard.edu, 651.270.9984.

Conflicts of interest: None.

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

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 ten 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 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 due to 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 live-born infants. Last menstrual period (LMP) was estimated using a validated algorithm based on the delivery data combined with diagnostic codes indicative of pre-term 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 three months prior to LMP, throughout pregnancy, and one month post-delivery.

For the current study, we further restricted the analytic sample to the 28,493 women who, in the 90 days prior to LMP, had a depression diagnosis (indicated by 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=2,437) 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=7,796) 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 and all psychiatric comorbidities (inpatient or outpatient) and comedications 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 (1 vs. 0, 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 above) in the claims records during this window. Because multiple days were rare, we dichotomized this outcome as any versus none and two or more versus zero or one. We also considered 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, schizophrenia) and select physical comorbidities (cancer, diabetes, headache disorder, epilepsy). Comedications included: anxiolytics, antipsychotics, lithium, benzodiazepines, barbiturates, opioids, anticonvulsants, and other hypnotics. Finally, we considered class of AD treatment (selective serotonin reuptake inhibitors, serotonin and norepinephrine reuptake inhibitors, tricyclics, bupropion, and other) as well as prior deliberate self-harm. 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 pre-treatment covariates described above. We then adjusted for the PS (i.e., the predicted probability of treatment continuation) in our 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 pre-defined 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 above.

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 due to 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 plots²⁶ to assess whether imbalances in measured confounders would lead to more bias in an IV 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 FDA 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.

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, 481 (7.0%) and 123 (14.0%) had an inpatient stay with depression; 30 (0.4%) and 12 (1.4%) 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 (OR=2.2, 95%CI 2.0–2.4). Adjustment for pre-specified 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. Pre-specified 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 pre-specified 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).

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 discontinued AD therapy it is by definition unclear 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 (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. Below 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 colleagues²² found evidence suggesting discontinuing AD therapy led to relapse, while Yonkers and colleagues³² 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 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 decisions to discontinue AD therapy were more frequently made without the physician's knowledge or 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 FDA 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 healthcare providers poses another potential source of confounding 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: 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. *American journal of obstetrics and gynecology*. Jun; 2007 196(6):544, e541–545. [PubMed: 17547888]
2. Andrade SE, Raebel MA, Brown J, et al. Use of antidepressant medications during pregnancy: a multisite study. *American journal of obstetrics and gynecology*. Feb; 2008 198(2):194, e191–195. [PubMed: 17905176]
3. Huybrechts KF, Palmsten K, Mogun H, et al. National trends in antidepressant medication treatment among publicly insured pregnant women. *General hospital psychiatry*. May-Jun;2013 35(3):265–271. [PubMed: 23374897]
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. [PubMed: 19739149]
5. Palmsten K, Setoguchi S, Margulis AV, Patrick AR, Hernandez-Diaz S. Elevated risk of preeclampsia in pregnant women with depression: depression or antidepressants? *American journal of epidemiology*. May 15; 2012 175(10):988–997. [PubMed: 22442287]
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. [PubMed: 22627901]
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. [PubMed: 23965506]
8. Huybrechts KF, Palmsten K, Avorn J, et al. Antidepressant use in pregnancy and the risk of cardiac defects. *The New England journal of medicine*. Jun 19; 2014 370(25):2397–2407. [PubMed: 24941178]
9. Walker AM. Confounding by indication. *Epidemiology*. Jul; 1996 7(4):335–336. [PubMed: 8793355]
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.
12. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 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. [PubMed: 16755261]
14. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Community Statistics*. 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. *Statistical methods in medical research*. Feb; 2013 22(1):70–96. [PubMed: 22016461]
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. [PubMed: 19770621]
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. [PubMed: 23840692]
18. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiology and drug safety*. Jan; 2013 22(1):16–24. [PubMed: 22550030]
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. [PubMed: 24146116]

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 : the journal of the American Medical Association*. Jun 18; 2003 289(23):3095–3105. [PubMed: 12813115]
21. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA : the journal of the American Medical Association*. Jan 6; 2010 303(1):47–53. [PubMed: 20051569]
22. Cohen LS, Altshuler LL, Harlow BL, et al. Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. *JAMA : the journal of the American Medical Association*. Feb 1; 2006 295(5):499–507. [PubMed: 16449615]
23. Miller M, Swanson SA, Azrael D, Pate V, Sturmer T. Antidepressant dose, age, and the risk of deliberate self-harm. *JAMA internal medicine*. Jun; 2014 174(6):899–909. [PubMed: 24782035]
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. [PubMed: 19487948]
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. *Journal of the American Statistical Association*. 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. [PubMed: 23549180]
28. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *The international journal of biostatistics*. 2007; 3(1):Article 14. [PubMed: 19655038]
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. [PubMed: 21163621]
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. [PubMed: 23532055]
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. [PubMed: 22872113]
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. [PubMed: 21900825]
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. *The Journal of clinical psychiatry*. Nov; 2011 72(11):e1563–1569. [PubMed: 22127200]
34. Sit D, Perel JM, Luther JF, Wisniewski SR, Helsel JC, Wisner KL. Disposition of chiral and racemic fluoxetine and norfluoxetine across childbearing. *Journal of clinical psychopharmacology*. Aug; 2010 30(4):381–386. [PubMed: 20631556]
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. *The Journal of clinical psychiatry*. 2006; 67(Suppl 4):27–30. [PubMed: 16683860]
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. [PubMed: 19671933]

Appendix: Definition and Evaluation of Proposed Instruments

Previous studies using preference-based instruments approximated provider's preference between treatment options as the treatment given to the provider's prior patient who met the

eligibility criteria (e.g., had the indication of interest). We adapted this to the current study by linking the available prescriber identification across prescription fills and patients (available in N=7,478). However, for patients who discontinued treatment, the only prescriber information we have available is that associated with their AD prescription fill prior to LMP (while for patients who continued AD therapy, we have prescriber information associated with AD prescription fills prior to LMP and during the first trimester). For consistency, we defined the relevant prescriber (and that prescriber's preference) based on AD prescriptions just prior to LMP for each patient. Supplemental Table 2 illustrates some limitations with this (and alternative) definitions.

We proposed three calendar-time-based instruments defined by an indicator for whether the eligible women had an LMP in the six months before versus the six months after each of the three FDA warnings regarding AD use issued during our study period (June 2003, December 2005, and July 2006). In June 2003, the FDA released a warning of increased risk for suicidal behaviors in youth associated with paroxetine use (this was later expanded to a black-box warning for other AD medications first for those under age 18 years and then later under 25 years). In December 2005, the FDA released a warning concerning use of paroxetine during pregnancy being related to an increased risk of heart defects in newborns. In July 2006, the FDA released a warning concerning the use of SSRIs during pregnancy being related to persistent pulmonary hypertension.

For geographic variation, we estimated the state-level proportion of women in our study who continued versus discontinued AD therapy, and dichotomized this as whether the women resided in a state with a high continuing proportion or a low continuing proportion (at or above 12.5% versus less than 12.5%). Other cut-offs for geography-variation-based proposed instruments were explored as sensitivity analyses, including those with exclusions to strengthen the proposed instrument (e.g., at or above 15% versus less than 10%, excluding those in states with 10–15% continuing proportions), but all assessed variants had similar strength and similar imbalances of covariates.

IV conditions (2) and (3) cannot be verified, and therefore require subject matter knowledge to justify their reasonableness. To try to justify these conditions, investigators often assess the distribution of measured confounders across levels of the proposed instrument: any imbalances found in measured confounders could be adjusted for, but may be indicative of similar or greater imbalances in unmeasured confounders. We used bias component plots to assess whether imbalances in measured confounders would lead to more bias in an IV versus non-IV analysis. Bias component plots present covariate balance (i.e., prevalence differences) of measured covariates across levels of the treatment and proposed instrument; however, they scale the covariate balance for the proposed instrument by its strength in order to appropriately account for the known bias amplification property of IV analyses. Specifically, the bias due to omitting a covariate in a non-IV analysis is a function of:

- i. the association between the covariate and treatment, and
- ii. the association between the covariate and outcome within levels of treatment.

The bias due to omitting a covariate in an IV analysis is a function of:

- (i') the association between the covariate and proposed instrument,
- (ii') the association between the covariate and outcome within levels of treatment, and
- (iii') the association between the proposed instrument and treatment.

Because (ii) and (ii') are the same, we can compare the non-shared bias components, i.e., (i) versus (i') and (iii'). The bias component plots present these non-shared components. See Jackson and Swanson (in press) for more details.

Figure 1.

Bias component plots for three proposed instruments

Plots display the non-shared bias components in an IV (i.e., by scaled IV) and non-IV analysis (i.e., by treatment) for each covariate. 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.

*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.

Table 1

Pre-LMP covariate distribution by AD 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=24103)	Continued AD (N=4390)	Discontinued AD (N=2033)	Continued AD (N=404)	Discontinued AD (N=6920)	Continued AD (N=876)
Demographics						
White	16882 (70.0)	3515 (80.1)	1315 (64.68)	322 (79.70)	4767 (68.89)	705 (80.48)
Age						
12–17	2205 (9.2)	274 (6.2)	294 (14.5)	39 (9.7)	793 (11.46)	86 (9.82)
18–24	10931 (45.4)	1360 (31.0)	944 (46.4)	134 (33.2)	3355 (48.48)	317 (36.19)
25–34	9211 (38.2)	2146 (48.9)	683 (33.6)	182 (45.1)	2340 (33.82)	375 (42.81)
35+	1756 (7.3)	610 (13.9)	112 (5.5)	49 (12.1)	432 (6.24)	98 (11.19)
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 Abuse	2231 (9.3)	455 (10.4)	872 (42.9)	183 (45.3)	NA	NA
Drug Abuse	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	18164 (75.4)	3312 (75.4)	1563 (76.9)	317 (78.5)	6920 (100.0)	876 (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)	13778 (57.2)	2357 (53.7)	1339 (65.9)	289 (71.5)	4417 (63.8)	540 (61.6)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Full Cohort		High Risk Subcohort		Low Risk Subcohort	
	Discontinued AD (N=24103)	Continued AD (N=4390)	Discontinued AD (N=2033)	Continued AD (N=404)	Discontinued AD (N=6920)	Continued AD (N=876)
Major Depressive Disorder Single Episode (296.2)	3597 (14.9)	661 (15.1)	683 (33.6)	107 (26.5)	802 (11.6)	126 (14.4)
Major Depressive Disorder Recurrent Episode (296.3)	6450 (26.8)	1468 (33.4)	866 (42.6)	184 (45.5)	1181 (17.1)	199 (22.7)
Other and Unspecified Episodic Mood Disorder (296.9)	80 (0.3)	11 (0.3)	13 (0.6)	NA	12 (0.2)	NA
Dysthymia (300.4)	3360 (13.9)	637 (14.5)	302 (14.9)	78 (19.3)	932 (13.5)	95 (10.8)
Adjustment Disorder with Depressed Mood (309.0)	1021 (4.2)	129 (2.9)	140 (6.9)	18 (4.5)	299 (4.3)	24 (2.7)
Prolonged Depressive Reaction (309.1)	106 (0.4)	12 (0.3)	NA	NA	36 (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)	471 (6.8)	51 (5.8)
Non-Psychiatric Comorbidities						
Cancer	582 (2.4)	133 (3.0)	70 (3.4)	19 (4.7)	109 (1.6)	19 (2.2)
Diabetes	412 (1.7)	106 (2.4)	66 (3.2)	22 (5.4)	83 (1.2)	14 (1.6)
Headache Disorder	1124 (4.7)	258 (5.9)	140 (6.9)	40 (9.9)	143 (2.1)	25 (2.9)
Epilepsy	122 (0.5)	42 (1.0)	30 (1.5)	13 (3.2)	11 (0.2)	NA

NA: cell counts <11 are suppressed based on CMS requirements.

Table 2

Propensity-score adjusted odds ratios for antidepressant continuation predicting inpatient depression days 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 ¹	2.1 (1.9, 2.3)	2.1 (1.6, 2.7)	1.5 (0.7, 3.1)
Model 2 ²	2.1 (1.9, 2.3)	2.0 (1.6, 2.6)	1.3 (0.7, 2.7)
Model 3 ³	2.0 (1.8, 2.2)	1.8 (1.4, 2.3)	1.2 (0.6, 2.6)
Model 4 ⁴	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

¹ Adjusts for age, race

² Adjusts for age, race, psychiatric comorbidities

³ Adjusts for age, race, psychiatric comorbidities, class of antidepressant, specific depression diagnosis

⁴ Adjusts for age, race, psychiatric comorbidities, class of antidepressant, specific depression diagnosis, other medications, other comorbidities, 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)	3.2 (1.6, 6.3)
Model 1 ¹	1.8 (1.4, 2.4)	1.6 (1.0, 2.7)	2.2 (1.8, 2.7)	3.2 (1.6, 6.3)
Model 2 ²	1.8 (1.4, 2.4)	1.6 (1.0, 2.7)	2.2 (1.8, 2.7)	3.2 (1.6, 6.3)
Model 3 ³	1.8 (1.3, 2.3)	1.4 (0.9, 2.3)	2.1 (1.7, 2.6)	3.0 (1.5, 6.0)
Model 4 ⁴	1.7 (1.3, 2.2)	1.2 (0.7, 2.0)	2.1 (1.7, 2.6)	3.0 (1.5, 5.9)

¹ Adjusts for age, race

² Adjusts for age, race, psychiatric comorbidities in the high risk subcohort. For low risk subcohort, this is the same as Model 1.

³ Adjusts for age, race, psychiatric comorbidities, class of antidepressant, specific depression diagnosis. For low risk subcohort, adjusts for age, race, and specific depression diagnosis.

⁴ 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.