



OPEN ACCESS

An electronic health record driven algorithm to identify incident antidepressant medication users

William V Bobo,¹ Jyotishman Pathak,² Hilal Maradit Kremers,³ Barbara P Yawn,⁴ Scott M Brue,⁵ Cynthia J Stoppel,¹ Paul E Croarkin,¹ Jennifer St Sauver,³ Mark A Frye,¹ Walter A Rocca^{3,6}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/amiajnl-2014-002699>).

¹Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA

²Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

³Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

⁴Division of Research, Olmsted Medical Center, Rochester, Minnesota, USA

⁵Biomedical Informatics Support System, Mayo Clinic, Rochester, Minnesota, USA

⁶Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA

Correspondence to

Dr William V Bobo, Department of Psychiatry and Psychology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA; bobo.william@mayo.edu

Received 31 January 2014

Revised 16 April 2014

Accepted 18 April 2014

Published Online First

29 April 2014



Open Access
Scan to access more
free content



CrossMark

To cite: Bobo WV, Pathak J, Kremers HM, et al. *J Am Med Inform Assoc* 2014;**21**:785–791.

ABSTRACT

Objective We validated an algorithm designed to identify new or prevalent users of antidepressant medications via population-based drug prescription records.

Patients and methods We obtained population-based drug prescription records for the entire Olmsted County, Minnesota, population from 2011 to 2012 (N=149 629) using the existing electronic medical records linkage infrastructure of the Rochester Epidemiology Project (REP). We selected electronically a random sample of 200 new antidepressant users stratified by age and sex. The algorithm required the exclusion of antidepressant use in the 6 months preceding the date of the first qualifying antidepressant prescription (index date). Medical records were manually reviewed and adjudicated to calculate the positive predictive value (PPV). We also manually reviewed the records of a random sample of 200 antihistamine users who did not meet the case definition of new antidepressant user to estimate the negative predictive value (NPV).

Results 161 of the 198 subjects electronically identified as new antidepressant users were confirmed by manual record review (PPV 81.3%). Restricting the definition of new users to subjects who were prescribed typical starting doses of each agent for treating major depression in non-geriatric adults resulted in an increase in the PPV (90.9%). Extending the time windows with no antidepressant use preceding the index date resulted in only modest increases in PPV. The manual abstraction of medical records of 200 antihistamine users yielded an NPV of 98.5%.

Conclusions Our study confirms that REP prescription records can be used to identify prevalent and incident users of antidepressants in the Olmsted County, Minnesota, population.

BACKGROUND AND SIGNIFICANCE

In the last 30 years, there have been pronounced increases in the use of antidepressants,^{1–3} which are now the third most commonly prescribed medication class in the USA.⁴ Although newer antidepressants (beginning with the selective serotonin reuptake inhibitors) are effective and generally well-tolerated for treating a number of psychiatric disorders,^{5–10} the explosive growth in the use of antidepressants continues to raise questions related to their effectiveness, side effects, and cost.^{11–13}

Clinical trials have been adequate to establish antidepressant efficacy and to quantify the incidence of common treatment-emergent adverse

effects; however, these data are limited by poor external validity.¹⁴ Large, randomized, comparative effectiveness studies may partially overcome this limitation,¹⁵ but they are unlikely to have sufficient power or duration of follow-up needed to detect outcomes that are infrequent or require a long time to develop.¹⁶ Furthermore, such trials typically exclude vulnerable but clinically important populations in which antidepressant use may be substantial (such as pregnant women and medically ill persons), and cannot address questions about antidepressant prescribing and use at the population level. Thus, many important questions related to antidepressant use, safety, and effectiveness must be evaluated outside of clinical trials using rigorously designed observational studies.¹⁷

The medical records linkage system maintained by the Rochester Epidemiology Project (REP) contains data on essentially all sources of medical care available to and utilized by the Olmsted County, Minnesota, population,¹⁸ and is therefore a potentially valuable resource for studying the use and effects of medications in a well-defined population. REP records of inpatient and outpatient medical care encounters allow identification of a phenotypically well-defined cohort, longitudinal follow-up of cohort members, and ascertainment of exposures, endpoints, and confounding variables.¹⁹ In particular, REP prescription records may provide objective, detailed, reliable, and relatively low-cost measures of drug exposure for large numbers of individuals that are not subject to recall bias (similar to other computerized prescription records²⁰), or systematic exclusion based on socioeconomic or clinical factors.²¹

However, because these records were collected during routine medical care and not for research purposes, they are subject to misclassification.²² In pharmacoepidemiologic studies, exposure misclassification can introduce biases that cannot be overcome using statistical adjustment or other data analytic techniques. Other authors have shown high concordance between computerized prescription records and patient self-report of medication use^{23–26} or medical record review.^{23 27–30} However, validation of REP prescription records of antidepressants or other medication exposures has not been performed.

In addition to minimizing the risk of misclassification of medication exposures, the ability to identify incident (new) users of antidepressants and other medications is crucial for conducting pharmacoepidemiologic studies of drug effectiveness or safety. Studies of prevalent users of study

medications may underestimate effects related to drug initiation, particularly effects that occur early in the course of treatment. We therefore conducted a study to validate a computer algorithm designed to identify new antidepressant users in automated REP antidepressant prescription records, using manual medical records review as a gold standard.

PATIENTS AND METHODS

Study population

The study sample was drawn from the full enumeration of all individuals residing in Olmsted County, Minnesota, between January 1, 2011, and December 31, 2012, identified using the REP census³¹ (n=149 629). All individuals who had given permission for their medical records to be used for research and were aged 6 years or more on the date of their first qualifying study drug prescription (index date) were considered eligible for the study.

Algorithm definition: antidepressant new users

We developed a computer algorithm designed to identify all potential new users of antidepressant medications (see online supplementary table S1) in the population during the study period. The list of study drugs in online supplementary table S1 represents all antidepressant drugs approved for clinical use in the USA during the study period. The algorithm required the exclusion of antidepressant prescriptions during the 6 months preceding the index date, including absence of drugs that were prescribed more than 6 months before the index date but were continued during the 6-month period. The estimated days of exposure for a given antidepressant that was prescribed preceding the 6-month period was determined using the date, days of supply, quantity dispensed, and number of refills for each prescription.

From the total pool of new antidepressant users defined by our algorithm, we selected a random sample of 200 subjects stratified by age group (<18 years, 18–64 years, and ≥65 years on the index date) and sex. We also selected an age- and sex-stratified random sample of 200 potential new users of antihistamine medications (see online supplementary table S1) drawn from persons who did not meet the computer definition of new antidepressant user. The algorithm to define new antihistamine users required the exclusion of both antihistamine and antidepressant exposure in the 6 months preceding the first qualifying antihistamine prescription. This sample was used to estimate true-negative and false-negative rates, and the negative predictive value (NPV).

Data resource and drug prescription data

Based on our previous work,³² we retrieved outpatient drug information from automated prescription systems used by the Mayo Clinic and Olmsted Medical Center (figure 1), which together provide most of the medical care for Olmsted County residents.^{32 33} More specifically, the prescription data from Olmsted Medical Center and Mayo Clinic were initially extracted using structured querying and natural language processing techniques, respectively, using Mayo's open-source cTAKES NLP platform.^{32 34} These data were ultimately mapped to RxNorm codes (<https://www.nlm.nih.gov/research/umls/rxnorm>) and grouped hierarchically according to NDF-RT categories by leveraging National Library of Medicine's RxNav and NDF-RT web service API calls.³² This step allowed for classification of the medication data using NDF-RT's legacy VHA Drug as well as Pharmacologic classes. Additional details about the mapping process are described in Pathak *et al.*³²

The current REP infrastructure employs an automated drug repository that retrieves electronic prescriptions from both of these systems via the methodology described above.³² Contained within these integrated electronic prescription files are data on drug name, form (eg, tablet, capsule, liquid, injection), dosage/strength, dosage frequency, quantity prescribed, date prescribed, and number of refills for each prescription. The repository also contains medication reconciliation data for all active drugs (prescribed or self-reported).^{35 36} Importantly, the structure of these data elements allowed us to perform structured queries by individual study drug names to identify potential new users of antidepressants in our cohort.

Medical records review and case adjudication

We sought to validate REP automated prescription data using manual medical record review as the standard for comparison. A trained study coordinator reviewed the records of all pertinent medical care for new users of antidepressants and antihistamines within 2 years (730 days) preceding the index date. We abstracted from the medical records the date of a medical encounter that mentioned initiation of a study drug or an indication of the actual date a study drug prescription was provided, key prescription data elements (drug name, dose, formulation, quantity dispensed, number of refills), study drug prescriptions preceding the index date, specialty of the provider who wrote the prescription, and indication for the prescription. All study data were recorded on de-identified abstraction sheets. The study was approved by the Institutional Review Boards at Mayo Clinic and Olmsted Medical Center.

Validated cases (true positives) were those that met the algorithm definition for antidepressant new user, with medical record review verification of the correct medication name, dose, formulation, number dispensed, and number of refills. Cases with an antidepressant prescription or overlapping period of antidepressant exposure within 6 months of the index date (prevalent antidepressant users), or those for which the index date or key prescription data could not be verified, were classified as false positives.

Statistical analyses

Clinical and demographic characteristics were presented as number (percentage) and median (inter-quartile range), as appropriate. Proportions were compared using χ^2 statistics. The positive predictive value (PPV) and NPV of the computer case definition for new antidepressant user was calculated with 95% CIs for binomial proportions using Wilson's formula. Case confirmation from manual medical record review served as the gold standard. All analyses were conducted using STATA statistical software, V.10.0 (STATA, College Station, Texas, USA).

We performed three sets of alternative analyses. First, we hypothesized that new antidepressant prescriptions would not exceed the starting doses of the agents of interest, and that antidepressant doses exceeding typical starting doses on the index date may represent prevalent use rather than new use. Thus, we calculated a standardized dose ratio for each antidepressant drug by dividing the antidepressant dose on the index date by the starting dose of that agent for treating major depressive disorder or other approved indications (eg, obsessive-compulsive disorder for clomipramine) in non-geriatric adults (see online supplementary table S2). We then recalculated PPV after excluding cases for which the standardized antidepressant dose ratio on the index date exceeded a value of 1.0. We recalculated PPV estimates after expanding the time window used to define new antidepressant use from 6 months to 1, 1.5, or 2 years preceding

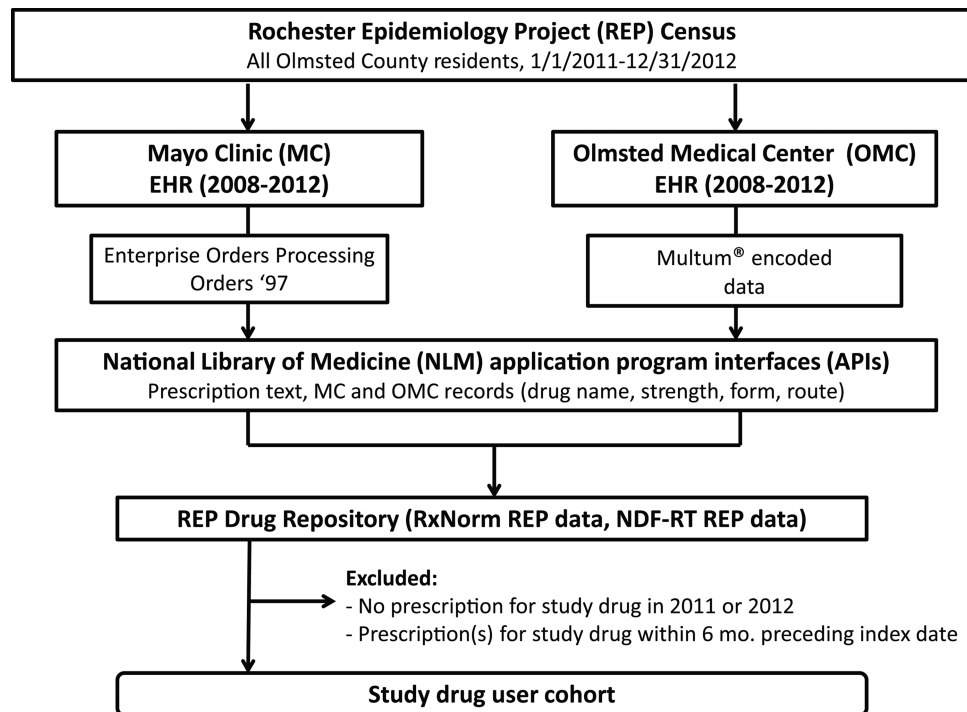


Figure 1 Schematic informatics workflow. The population of Olmsted County, Minnesota (January 1, 2011–December 31, 2012) was obtained using the Rochester Epidemiology Project (REP) census. For these patients, structured medication data from Mayo Clinic (MC) and Olmsted Medical Center (OMC) electronic health record (EHR) systems was obtained using Enterprise Orders Processing (from 2009 onward), Orders '97 (prior to 2009), and a commercial drug database (Multum) via a series of SQL queries. The collective data were then represented as RxNorm encoded data that were further categorized under NDF-RT drug classes, using the National Library of Medicine's NDF-RT web services API. To create drug user cohorts from which random samples for this study were drawn, individuals were excluded if they had no prescription for a study drug in 2011 or 2012 and if there was evidence of a study drug prescription in the 6 months preceding the index date (ie, the date of the first qualifying study drug prescription).

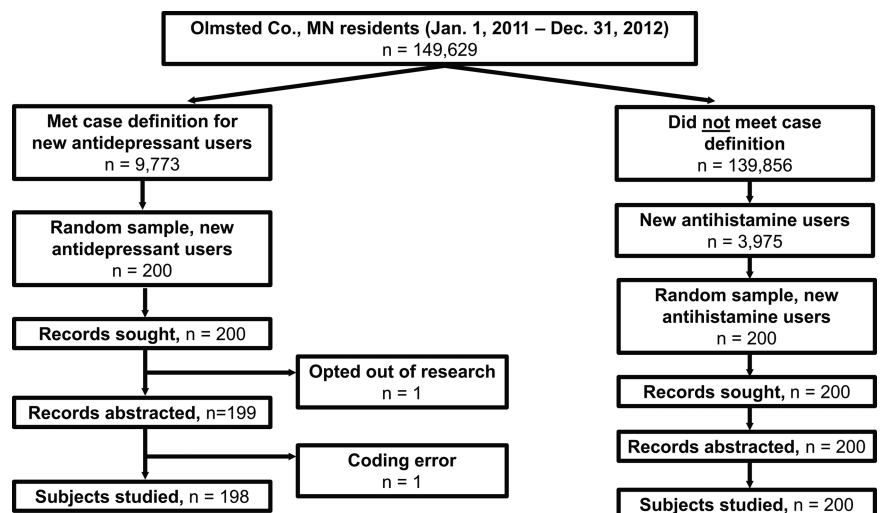
the index date. We hypothesized that the expansion would result in incrementally higher PPVs for new antidepressant use, with or without exclusion of cases based on a standardized antidepressant dose ratio >1.0 . We hypothesized that the misclassification rate of prevalent antidepressant use as new use would decrease with increasing duration of residence in Olmsted County, Minnesota, prior to the index date. Therefore, we recalculated PPV estimates after restricting the sample to subjects who resided in Olmsted County, Minnesota, for at least 1 year preceding the index date.

RESULTS

Cohort demography and characteristics

A total of 9773 (6.5% of 149 629) subjects met the computer case definition of antidepressant new user, from which 200 cases were selected at random, with sampling stratified by age and sex. Medical records were not reviewed or adjudicated for one subject because of a change in authorization for use of medical records for research and one subject because of a data coding error (figure 2). This left a total of 198 adjudicated potential antidepressant new users. The sample was

Figure 2 Sample for validation of computer definition of antidepressant new user. All residents of Olmsted County, Minnesota, who had given permission for their medical records to be used for research and were ≥ 6 years of age between January 1, 2011 and December 31, 2012 were considered eligible for this study. The study sample consisted of 198 subjects electronically identified as new antidepressant users, as detected by our computer algorithm. A random sample of 200 subjects were electronically identified as new antihistamine users from the total pool of subjects who did not meet our computer algorithm definition of new antidepressant user.



predominantly Caucasian and middle-aged, with an approximately equal sex distribution (table 1). Most antidepressant use involved selective serotonin reuptake inhibitors.

Overall performance of computer definition of new user

Of the 198 new users, 161 were confirmed as new users after manual medical records review, resulting in a PPV of 81.3% (95% CI 76.4% to 87.0%). A total of 37 cases were misclassified, 36 of which were due to misclassifying prevalent users as new users, that is, evidence of antidepressant use was identified within 6 months of the index date based on review of medical records (table 2). We could not confirm new or prevalent use of an antidepressant in the medical records of one case. Thus, the PPV for prevalent use of antidepressants on the index dates was 99.5% (197/198, 95% CI 97.2% to 99.9%). PPV estimates did not vary significantly by age stratum or sex, although numerically lower PPV estimates were observed for males relative to females, and persons aged 65+ years relative to those aged 18–64 and those under the age of 18 years (table 2). PPV estimates also did not differ significantly by general indication (psychiatric, non-psychiatric) or antidepressant prescriber specialty (non-physician, primary care physician, psychiatrist, non-psychiatric specialty physician; data not shown).

Computer definition performance: alternative definitions

After restricting the sample to cases in which the standardized antidepressant dose ratio was ≤ 1.0 on the index date (n=165), 150 cases were confirmed as new users (PPV 90.9%, 95% CI 85.5% to 94.4%) (table 2). All 15 misclassified cases were prevalent antidepressant users. Extending the time windows preceding the index date to 1, 1.5, or 2 years resulted in incrementally greater reductions in the number of potential new antidepressant users, with only modest increases in the PPV

(table 2). Expanding the time windows preceding the index date and restricting the cohort to cases with standardized antidepressant dose ratios ≤ 1.0 on the index date resulted in PPV estimates of 90.8%. Requiring at least one continuous year of Olmsted County, Minnesota, residence prior to the index date also did not improve the PPV estimate (table 2).

Subjects who did not meet the computer definition: NPV and sensitivity estimates

Of the 139 856 subjects who did not meet the algorithm definition of new antidepressant user, a total of 3975 were classified as potential new antihistamine users. From these subjects, we derived a random sample of 200 users stratified by age and sex, in a manner identical to that used to define the random sample of potential new antidepressant users. Medical records were reviewed manually for all 200 cases, three of whom had evidence of antidepressant use in the 6 months preceding the first qualifying antihistamine prescription, yielding a false negative rate of 0.015 (or 1.5%) and NPV estimate of 98.5% (95% CI 95.7% to 99.5%). Our sampling scheme was based on first identifying a random sample of new antidepressant user and non-user cases based on the computer algorithm. As such, we were unable to directly calculate algorithm sensitivity, which would require a random sample of the entire population subsequently classified as antidepressant new users and non-users according to manual medical record review. However, as a secondary approach, and making a number of assumptions, we have provided a crude estimate of algorithm sensitivity based on data from our study (see online supplementary appendix).

DISCUSSION

We validated an algorithm for identifying new antidepressant users in REP prescription records using manual review of medical records as a gold standard. In the sample studied, the PPV was 81.3%. Restricting the sample to cases in which the standardized antidepressant dose ratio on the index date was ≤ 1.0 resulted in a marked increase in the PPV (90.9%), whereas extending the time windows preceding the index date to 1, 1.5, or 2 years resulted in only modest improvements in the PPV at the expense of incrementally greater loss of potential new antidepressant users. Requiring a minimum of 1 year of Olmsted County residence did not increase the PPV estimates.

Over the last 40 years, the REP has successfully provided the facilities and data for over 2000 descriptive, case-control, and cohort studies.¹⁸ Strengths of the REP are well-documented, and include the ability to link nearly all medical records and other sources of healthcare data for the entire population of Olmsted County, Minnesota. Only recently, however, has the REP medical records linkage and indexing system also included detailed and complete automated drug prescription information.^{19–33} Historically, drug prescription information was available in the original medical records, but intensive efforts and expense on the part of investigators were required to manually abstract pertinent data related to individual drug prescriptions. Therefore, the new electronic prescription records may provide reliable, efficient, and low-cost data suitable for population-based studies of medications.²⁰

However, there are several challenges involved with the use of automated prescription data. Because REP and other computerized prescription records are typically collected during routine medical care and not for research purposes, they are subject to misclassification.^{16–22} Prior studies have investigated the concordance between computerized pharmacy records and patient self-report data^{23–26} or review of medical records.^{23–27–30} Nearly all of these

Table 1 Demographic and clinical characteristics of the study sample

	Antidepressant users N=198		Antihistamine users* N=200	
	Median	IQR	Median	IQR
Age (years)	44.5	17–65	41.5	22.5–57.5
	N	%	N	%
Gender				
Male	98	49.5	99	49.5
Female	100	50.5	101	50.5
Race/ethnicity				
Caucasian	187	94.4	145	72.5
African-American	3	1.5	19	9.5
Hispanic	1	0.5	4	2.0
Asian	3	1.5	17	8.5
Other/unknown	5	2.0	15	7.5
Antidepressant class				
SSRI	116	58.6	–	–
SNRI	21	10.6	–	–
TCA	30	15.2	–	–
MAOI	0	0.0	–	–
Other	31	15.6	–	–

*These subjects were randomly selected among all non-users of an antidepressant drug in 2011–2012.
MAOI, monoamine oxidase inhibitor; SNRI, serotonin-norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

Table 2 Comparison of electronically defined new users of an antidepressant medication versus manual review of medical records

Classification	Electronically defined (N)	Manual medical records abstraction (N)			PPV (95% CI)
		New AD user	Prevalent AD user	No AD prescription	
Total	198	161	36	1	81.3 (76.4 to 87.0)
Standardized dose ratio ≤ 1.0	165	150	15	0	90.9 (85.5 to 94.4)
Expanded time windows* (years)					
1	180	147	32	1	81.7 (75.4 to 86.6)
1.5	167	138	28	1	82.6 (76.2 to 87.6)
2	160	132	27	1	82.5 (75.9 to 87.6)
Olmsted County residence >1 year†	192	155	36	1	81.3 (74.6 to 85.7)
By sex‡					
Male	98	77	21	0	78.6 (69.5 to 85.5)
Female	100	84	15	1	84.0 (75.6 to 89.9)
By age strata§ (years)					
<18	55	47	8	0	85.5 (73.8 to 92.4)
18–64	93	76	17	0	81.7 (72.7 to 88.3)
65+	50	38	11	1	76.0 (62.6 to 85.7)

*Time periods (1, 1.5, or 2 years) refer to the period of days immediately preceding the index date for which there was no evidence of antidepressant exposure. Further restriction to participants with standardized antidepressant dose ratio ≤ 1.0 improved PPV estimates to 90.8% (95% CI 85.1% to 94.4%) for the 1-year window, to 90.8% (95% CI 85.0% to 94.6%) for the 1.5-year window, and to 90.8% (95% CI 85.0% to 94.6%) for the 2-year window.

†Requiring a minimum of 1 year of Olmsted County residence prior to the index date and further restriction to participants with standardized antidepressant dose ratio ≤ 1.0 improved the PPV estimate to 90.6% (95% CI 85.1% to 94.2%).

‡Restriction to participants with standardized antidepressant dose ratio ≤ 1.0 improved the PPV estimate to 89.5% (95% CI 80.6% to 94.6%) for males and to 93.2% (95% CI 85.9% to 96.8%) for females.

§Restriction to participants with standardized antidepressant dose ratio ≤ 1.0 improved the PPV estimate to 93.9% (95% CI 83.5% to 97.9%) for subjects aged <18 years, to 92.0% (95% CI 83.6% to 96.3%) for subjects aged 18–64 years, and to 85.4% (95% CI 71.6% to 93.1%) for subjects aged 65+ years.

AD, antidepressant; PPV, positive predictive value.

studies focused on validation of prescription records for elderly persons who were health maintenance organization or Medicaid beneficiaries. Few studies focused on validation of antidepressants. McKenzie *et al*²⁷ validated Oregon Medicaid pharmacy claims data for antidepressants and other psychotropic drugs in a cohort of 900 adult nursing home residents using medical chart review as a gold standard. The PPV for antidepressant exposure was 93.8%, with a high correlation between the average daily antidepressant dose in the pharmacy claims data and medical records ($r=0.81-0.84$, depending on the algorithm used). Shorr and colleagues validated Tennessee Medicaid pharmacy claims data for tricyclic antidepressant use against medication administration records for 9452 adult nursing home residents.²⁸ There was high concordance (85–88%) between Medicaid files and medical records with regard to antipsychotic drug use. Although no data were presented for antidepressants, the authors reported that the degree of agreement between Medicaid files and medical records for antidepressant use was similar to that for antipsychotic use. Finally, Johnson and Vollmer documented a PPV of 76% for automated outpatient prescription records of psychotropic medications (including antidepressants) using an in-home assessment as a gold standard for 83 frail elderly health maintenance organization enrollees.²⁵

Our study is the first systematic attempt at validating automated REP prescription data for any specific medication class. Using this data source, Zhong *et al*³³ previously reported that antidepressants were the second-most frequently prescribed drug group in the entire Olmsted County, Minnesota, population in 2009 ($n=18\,028$, 13%), and the most commonly prescribed drug class in the 30–40-year-old population ($n=6310/37\,927$, 17%). However, the Zhong *et al* study described prevalent use of antidepressants and did not focus on the start of new medication use (incident use). Our data confirmed the validity of REP prescription records for identifying both prevalent and incident antidepressant users.

Our validation efforts began with the goal of developing an algorithm for identifying new antidepressant users. Prevalent users have, by definition, persisted in their use of a medicine and have presumably tolerated and perceived benefit from it.^{17, 37} This can significantly impact the validity of study findings from mixed prevalent and incident user cohorts by underestimating effects that occur relatively soon after drug initiation, over-emphasizing the effects of long-term medication use, over-representing persons with good adherence, and potentially over-estimating favorable treatment outcomes.³⁸ Furthermore, studying drug effects in prevalent users permits adjustment only for potential confounding factors assessed after study drug exposure.¹⁷ Adjustment for these factors may lead to an under-estimation of study drug effect if the factors are influenced by study drug exposure and are on the causal pathway between the study drug and the outcome of interest.¹⁷

REP prescription data have several potentially important advantages. The REP data systems permit access to data on prescription records and health conditions for an entire geographically defined population, making it unique among data sources in the USA.³⁹ Because REP is based on a geographically defined population, it is not subject to healthy beneficiary bias that may be encountered in managed-care systems or referral bias that may be present in drug exposure registries.^{40, 41} REP data are not limited by socioeconomic factors (as is the case with Medicaid data), age (Medicare data), healthcare provider type, or insurance status.³⁹ Importantly, REP prescription data are linked to the complete medical records of each cohort member, thus enabling precise ascertainment of the intended antidepressant indications and future investigations of the immediate and delayed effects of antidepressants on a variety of health outcomes (including conditions or effects not adequately captured by diagnostic codes, such as changes in laboratory results or non-specific symptoms), as well as investigations linking prescription data with genetic variants (population-level pharmacogenetics studies) and cost data (cost–outcome studies).

The completeness and accuracy of data on medication exposure is known to vary across data sources,¹⁶ and these factors have a direct influence on the PPV and generalizability of our approach to other data sets. Concerning completeness, the REP infrastructure includes prescription data from outpatient settings (including office-based and hospital-based outpatient practices),³³ as well as information on self-reported use of over-the-counter medications and alternative therapies for some cohort members. Concerning accuracy, our results suggest that a high proportion of antidepressant new user cases identified by our algorithm will be true cases; however, a more acceptable level of predictive value was reached only after eliminating cases with standardized doses that exceed the usual starting doses of a given drug for specific indications. Of the 37 misclassified cases, 36 were due to the discovery of prevalent antidepressant use after medical record review. The generalizability of our approach to other drug classes and other data sources requires further investigation.

There are additional limitations to consider. First, our study focused on validating antidepressant exposures in a single random sample of Olmsted County, Minnesota, residents and involved the testing of an algorithm designed for use in REP prescription data. Second, although the Olmsted County population captured by the REP system is large, it may not be large enough for adequately powered studies of rare outcomes related to a given drug exposure. Third, our definition of new antidepressant prescribing was conservative, based on an assumption that all antidepressant prescriptions written prior to the index date were filled and the entire medication supply was used as directed. This could result in over-exclusion of new users misclassified as prevalent users. Fourth, we were unable to verify whether antidepressant prescriptions were filled or that the medicines were ingested. This is perhaps less of a concern for studies of prescribing behavior. However, these limitations are important for studies of outcomes associated with antidepressant exposure. Finally, our objective was to test an algorithm that would identify potential but not established new antidepressant users; therefore, our sampling scheme began with classification of new antidepressant use and antidepressant non-use based on the computer algorithm definition. We did not draw a random sample of cases from the underlying population and subsequently classify them as new antidepressant users or non-users based on medical record review. This approach permitted direct calculation of the PPV, consistent with our study objectives, but only an indirect calculation of sensitivity. However, we provided a crude estimate of sensitivity based on a random sample of new antihistamine users with no evidence of antidepressant exposure, after making a number of assumptions. The new antihistamine users represented a small proportion of the total number of persons in the population not meeting the definition of new antidepressant user (~3%). Although we do not expect the antihistamine users as a group to be substantially different than the entire population of patients not meeting the algorithm definition of new antidepressant user, our sensitivity estimate should be interpreted with caution.

CONCLUSIONS

In conclusion, we validated a computer algorithm designed to identify new users of antidepressants in comprehensive prescription records for a geographically well-defined population in the USA. The algorithm had an overall PPV of 81% that increased to 91% after exclusion of cases in which the standardized antidepressant dose ratio on the index date exceeded a value of 1.0. Our results confirm that REP prescription records can be used

to identify prevalent and incident users of antidepressants in this population.

Contributors WVB, HMK, and WAR designed the study. WVB wrote the manuscript. All other authors checked individual sections in the manuscript and the discussion section. All authors contributed to and approved the final manuscript.

Funding This work was supported in part by National Institutes of Mental Health grant MH087747. This study was made possible through use of the resources of the Rochester Epidemiology Project, which is supported by the National Institute on Aging (R01 AG034676; PI: Rocca).

Competing interests None.

Ethics approval Institutional Review Boards at Mayo Clinic and Olmsted Medical Center approved this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- 1 Pincus HA, Tanielian TL, Marcus SC, *et al.* Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. *JAMA* 1998;279:526–31.
- 2 Olsson MM, Marcus SC. National patterns in antidepressant medication treatment. *Arch Gen Psychiatry* 2009;66:848–56.
- 3 Mojtabai R, Olsson M. Proportion of antidepressants prescribed without a psychiatric diagnosis is growing. *Health Aff (Millwood)* 2011;30:1434–42.
- 4 Hsiao CJ, Cherry DK, Beatty PC, *et al.* National Ambulatory Medical Care Survey: 2007 summary. *Natl Health Stat Rep* 2010;3:1–32.
- 5 Fournier JC, DeRubeis RJ, Hollon SD, *et al.* Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010;303:47–53.
- 6 Levkovitz Y, Tedeschi E, Papakostas GI. Efficacy of antidepressants for dysthymia: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry* 2011;72:509–14.
- 7 Kapczynski F, Lima MS, Souza JS, *et al.* Antidepressants for generalized anxiety disorder. *Cochrane Database Syst Rev* 2003;(2):CD003592.
- 8 Bakker A, van Balkom AJ, Stein DJ. Evidence-based pharmacotherapy of panic disorder. *Int J Neuropsychopharmacol* 2005;8:473–82.
- 9 Stein DJ, Ipser JC, Seedat S. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2006;(1):CD002795.
- 10 de Menezes GB, Coutinho ES, Fontenelle LF, *et al.* Second-generation antidepressants in social anxiety disorder: meta-analysis of controlled clinical trials. *Psychopharmacology (Berl)* 2011;215:1–11.
- 11 Jureidini J, Tonkin A. Overuse of antidepressant drugs for the treatment of depression. *CNS Drugs* 2006;20:623–32.
- 12 Conti R, Busch AB, Cutler DM. Overuse of antidepressants in a nationally representative adult patient population in 2005. *Psychiatr Serv* 2011;62:720–6.
- 13 Chung S. Does the use of SSRIs reduce medical care utilization and expenditures? *J Ment Health Policy Econ* 2005;8:119–29.
- 14 Wisniewski S, Rush AJ, Nierenberg AA, *et al.* Can phase III results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am J Psychiatry* 2009;166:599–607.
- 15 Depp C, Lebowitz BD. Clinical trials: bridging the gap between efficacy and effectiveness. *Int Rev Psychiatry* 2007;19:531–9.
- 16 Ray WA. Population-based studies of adverse drug effects. *N Engl J Med* 2003;349:1592–4.
- 17 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158:915–20.
- 18 Rocca WA, Yawn BP, St Sauver JL, *et al.* History of the Rochester Epidemiology Project: half a century of medical records linkage in a United States population. *Mayo Clin Proc* 2012;87:1202–13.
- 19 St Sauver JL, Grossardt BR, Yawn BP, *et al.* Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epidemiol* 2012;41:1614–24.
- 20 Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989;129:837–49.
- 21 St Sauver JL, Grossardt BR, Leibson CL, *et al.* Generalizability of epidemiological findings and public health decisions: an illustration from the Rochester Epidemiology Project. *Mayo Clin Proc* 2012;87:151–60.
- 22 Bobo WV, Cooper WO, Stein CM, *et al.* Positive predictive value of a case definition for diabetes mellitus using automated administrative health data in children and youth exposed to antipsychotic drugs or control medications: a Tennessee Medicaid study. *BMC Med Res Methodol* 2012;12:128.

- 23 Landry JA, Smyer MA, Tubman JG, *et al.* Validation of two methods of data collection of self-reported medicine use among the elderly. *Gerontologist* 1988;28:672–6.
- 24 West SL, Savitz DA, Koch G, *et al.* Recall accuracy for prescription medications: self-report compared with database information. *Am J Epidemiol* 1995;142:1103–12.
- 25 Johnson RE, Vollmer WM. Comparing sources of drug data about the elderly. *J Am Geriatr Soc* 1991;39:1079–84.
- 26 Curtis JR, Westfall AO, Allison J, *et al.* Agreement and validity of pharmacy data versus self-report for use of osteoporosis medications among chronic glucocorticoid users. *Pharmacoepidemiol Drug Saf* 2006;15:710–18.
- 27 McKenzie DA, Semradek J, McFarland BH, *et al.* The validity of medicaid pharmacy claims for estimating drug use among elderly nursing home residents: the Oregon experience. *J Clin Epidemiol* 2000;53:1248–57.
- 28 Shorr RI, Fought RL, Ray WA. Changes in antipsychotic drug use in nursing homes during implementation of the OBRA-87 regulations. *JAMA* 1994;271:358–62.
- 29 Mangione-Smith R, Wong L, Elliott MN, *et al.* Measuring the quality of antibiotic prescribing for upper respiratory infections and bronchitis in 5 US health plans. *Arch Pediatr Adolesc Med* 2005;159:751–7.
- 30 Maselli JH, Gonzales R. Measuring antibiotic prescribing practices among ambulatory physicians: accuracy of administrative claims data. *J Clin Epidemiol* 2001;54:196–201.
- 31 St Sauver JL, Grossardt BR, Yawn BP, *et al.* Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. *Am J Epidemiol* 2011;173:1059–68.
- 32 Pathak J, Murphy SP, Willaert BN, *et al.* Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project. *AMIA Annu Symp Proc* 2011;2011:1089–98.
- 33 Zhong W, Maradit-Kremers H, St Sauver JL, *et al.* Age and sex patterns of drug prescribing in a defined American population. *Mayo Clin Proc* 2013;88:697–707.
- 34 Savova GK, Masanz JJ, Ogren PV, *et al.* Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17:507–13.
- 35 Varkey P, Resar RK. Medication reconciliation implementation in an academic center. *Am J Med Qual* 2006;21:293–5.
- 36 Varkey P, Cunningham J, Bisping DS. Improving medication reconciliation in the outpatient setting. *Jt Comm J Qual Patient Saf* 2007;33:286–92.
- 37 Johnson ES, Bartman BA, Briesacher BA, *et al.* The incident user design in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2013;22:1–6.
- 38 Cox E, Martin BC, Van ST, *et al.* Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value Health* 2009;12:1053–61.
- 39 St Sauver JL, Warner DO, Yawn BP, *et al.* Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. *Mayo Clin Proc* 2013;88:56–67.
- 40 Kremers HM, Myasoedova E, Crowson CS, *et al.* The Rochester Epidemiology Project: exploiting the capabilities for population-based research in rheumatic diseases. *Rheumatology (Oxford)* 2011;50:6–15.
- 41 Melton LJ III, Dyck PJ, Karnes JL, *et al.* Non-response bias in studies of diabetic complications: the Rochester Diabetic Neuropathy Study. *J Clin Epidemiol* 1993;46:341–8.