Postprint Version	1.0
Journal website	http://www3.interscience.wiley.com/cgi-bin/fulltext/121496945/PDFSTART
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/18980206
DOI	10.1002/pds.1677

An algorithm to identify antidepressant users with a diagnosis of depression from prescription data

HELGA GARDARSDOTTIR PHARMD[1], ANTOINE C. G. EGBERTS PHD[1,2], LISET VAN DIJK PHD[3], MIRIAM C. J. M. STURKENBOOM PHD[4] AND EIBERT R. HEERDINK PHD[1]*

1Division of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands
2Department of Clinical Pharmacy, University Medical Centre Utrecht, Utrecht, The Netherlands
3Netherlands Institute for Health Services Research (NIVEL), Utrecht, The Netherlands
4Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands

SUMMARY

Purpose Antidepressants are used for many indications besides depression. This makes investigating depression treatment outcomes in prescription databases problematic when the indication is unknown. The aim of our study is to develop an algorithm to identify antidepressant drug users from prescription data that suffer from depression.

Methods Data for deriving the algorithm were obtained from the Second Dutch National Survey of General Practice, carried out in 2001 by The Netherlands Institute for Health Services Research (NIVEL), and for validation the Integrated Primary Care Information (IPCI) database was used. Both sets included adults receiving their first antidepressant drug in 2001 (n¼1855 and 3321, respectively). The outcome was a registered diagnosis of depression. Covariates investigated for developing the algorithm were patient and prescribing characteristics, and comedication.

Results The predictive algorithm included age, SSRI prescribed on the index date, prescribed dose, general practitioner as prescriber and the number of antidepressant prescriptions prescribed plus medication for treating acid related disorders, laxatives, cardiac therapy or hypnotics/ sedatives prescribed in the 6 months prior to index date. The probability that the algorithm correctly identified an antidepressant drug user as having a depression diagnosis was 79% with a sensitivity of 79.6% and a specificity of 66.9%.

Conclusion In conclusion, we developed and validated an algorithm that can be used to compose cohorts of patients treated with antidepressants for depression from prescription databases. Copyright # 2008 John Wiley & Sons, Ltd.

INTRODUCTION

Many studies have been performed on antidepressant drug use investigating a multitude of outcomes such as efficacy and tolerability,1 patterns of prescribing,2–7 user characteristics8,9 and health utilization costs.10 Studies on outcomes of antidepressant drug use have often been performed in prescription databases and consequently the results focus on the outcomes of antidepressant drug use rather than that of the diseases and symptoms that the antidepressants are intended to treat. Although the name 'antidepressant' correctly suggests that these medicines can be used to treat depression, nowadays the indications for antidepressant drug prescribing have broadened. Antidepressants are prescribed for a variety of psychiatric illnesses such as generalized anxiety disorder, obsessive–compulsive disorder, panic disorder, social phobia and bulimia nervosa. In addition, antidepressants are prescribed for somatic illnesses such as enuresis/incontinence, sleeping disorders, migraine prophylaxis, functional dyspepsia and neuropathic pain.11–17 A recent study investigating the indications for which antidepressant drugs are prescribed in general practice revealed that general practitioners (GPs) prescribe antidepressants for various diseases and symptoms of which depression only contributed to around 45.0% while anxiety accounted for 17.0%, headache 2% and obsessive– compulsive disorder and phobias for 1.5%.18 Unfortunately most prescription databases do not include information on the clinical reasons for prescribing. Researchers have tried to solve this problem by developing disease specific algorithms.
19–22 The development of such an algorithm is performed in databases where information on patient characteristics, medication use and diagnoses are available. The developed algorithm can subsequently be used in prescription databases to calculate the patient's probability of having a specific disorder based on a covariate profile including the patient's characteristics, prescriber information and medication use.
Spettell et al.23 investigated simple algorithms to identify patients with depression from administrative data, using both diagnostic and prescription data, but to our knowledge no such algorithm has been developed for use in prescription databases where diagnostic data are not readily available.
The aim of our study is to develop and validate an algorithm that can be used to compose cohorts of patients treated with antidepressants for depression from prescription databases.

METHODS

A diagnostic study was performed to develop an algorithm that identifies patients that use an antidepressant to treat depression from prescription databases. The accuracy of the diagnostic algorithm was validated in a different population in The Netherlands.

Setting and study population

Data for the derivation set were obtained from the Second Dutch National Survey of General Practice (DNSGP-2) which was carried out in 2001 by The Netherlands

Institute for Health Services Research (NIVEL) and has been described in detail elsewhere.24 In short, 195 GPs in 104 practices registered all physician–patient contacts during 12 months. The GPs registered all health problems presented within a consultation in a standardized manner and diagnoses were coded using the International Classification of Primary Care (ICPC) scheme.25 A part of the patients included in the DNSGP-2 have been linked to pharmacy dispensing data, which includes prescription data for the patients from 1999 until the end of 2003.26 Thus, each patient had a complete prescription history spanning at least the 12 months prior to and 12 months post index date. The pharmacy dispensing data contains information on the dispensed drug, dispensing date, amount dispensed, prescribed dosage regimen and the estimated duration of use. Drugs were coded according to the Anatomical Therapeutic Chemical (ATC) classification.27 Patient information per prescribed medicine includes gender and date of birth. Each patient is identified with an anonymous unique patient-identification code that allows for the observation of patient medication use in time. Since virtually all patients in The Netherlands are registered with a single community pharmacy, independent of prescriber, pharmacy records are near complete with regard to prescription drugs.28 The source population included individuals registered in the DNSGP-2 who could be linked to pharmacy dispensing data (n¼110 078). Patients, 18 years and older, from the source population who got an antidepressant prescription dispensed from a pharmacy in the year 2001 where selected (n¼5140). The date of the first dispensed antidepressant prescription in year 2001 was set as index date.

Only new starters of antidepressant drug therapy (n¼1855) were selected thus those with any antidepressant drug use in the 12 months preceding the index date were excluded. In The Netherlands, the following antidepressants were available and prescribed during the study period: Tricyclic antidepressants; TCAs (amitriptyline, clomipramine, desipramine, dosulepin, doxepin, imipramine, maprotiline, nortriptyline, trimipramine), selective serotonin reuptake inhibitors; SSRIs (citalopram, fluoxetine, fluvoxamine, paroxetine, sertraline) and other (mianserin, mirtazapine, moclobemide, nefazodone, oxitriptan, phenelzine, trazodone, tranylcypromine, venlafaxine).

Data for the validation set were obtained from the Integrated Primary Care Information (IPCI) database at the department of Medical Informatics of the Erasmus Medical Center in Rotterdam, The Netherlands. The IPCI database is a general practice research database with coded and anonymous electronic patient records of more than 800 000 patients from approximately 100 GPs. The database includes information on demographics, symptoms and diagnoses (ICPC and free text), clinical and laboratory findings, referrals, hospitalizations and prescription information. Information on prescribed medicines includes drug name, ATC code, dose, dosage form, prescribed quantity and indication for prescribing.

The source population included individuals, 18 years and older, registered in the IPCI database in the year 2001 (n¼341 498). The study population was composed of patients from the source population who got an antidepressant prescription from a GP in 2001 (n¼7478). The date of the first dispensed antidepressant prescription in 2001 was set as index date. Only new starters of antidepressant drug therapy (n¼3321) were selected thus those with any antidepressant drug use in the 12 months preceding the index date were excluded. Each patient had a complete prescription history spanning the 12 months prior to and 12 months post index date.

Development of diagnostic algorithm

The outcome was defined as depression diagnosed and registered by a GP. The antidepressant drug users who at any point during the 12 months of the DNSPG-2 study had the ICPC codes 'depression' or 'feeling depressed' (ICPC¼P03, P76) registered in their medical file were defined as having a depression diagnosis. The antidepressant drug users that did not have a depression or feeling depressed diagnosis registered in their medical file had other diagnoses that could be associated with their antidepressant drug use such as anxiety disorder, obsessive–compulsive disorder, eating disorders, enuresis, incontinence, headache, migraine, sleeping disorder or neuropathic pain.

For about a third of the study population no registered or clinically accepted indication for antidepressant drug prescribing was found in the medical file.18 The same method was used to define the outcome for the validation set as for the derivation set. Antidepressant drug users were defined as having a depression diagnosis if the ICPC codes 'depression' or 'feeling depressed' (ICPC¼P03, P76) were found in their medical file as indication for prescribing or free text during the 12 months around the index date.

Covariates

The covariates that were investigated in this study were patient characteristics (age, gender), antidepressant characteristics [type of antidepressant prescribed at index date (TCA, SSRI, other), dose of the 2nd consequent antidepressant prescription following index date, if the first and second consequent antidepressant prescription differed (switch), number of antidepressant prescriptions received in 12 months following index date, type of prescriber at index date (GP, specialist/other/unknown)] and co-medication prescribed 6 months prior to and 6 months post index date. To investigate co-medication, all medicines were divided into groups according to second level ATC grouping, for example A01, A02, etc. Each second level ATC group was investigated as a covariate. We chose to select the dose of the 2nd antidepressant prescription as a covariate to compensate for gradual titration to effective therapeutic dose, as is custom in the beginning of an antidepressant drug treatment. The dose is expressed as amount of defined daily dose (DDD). A 2nd prescription was considered to be a consequent prescription if it was dispensed within 30 days following the theoretical end date (index date plus amount dispensed divided by dose) of the antidepressant drug prescription prescribed on index date. For those antidepressant drug users without a 2nd consequent prescription the dose for the antidepressant prescription on index date was used.

Data analysis

The association between a diagnosis of depression and the potential diagnostic covariates was quantified using univariate logistic regression analysis. The independent contributions of covariates with an univariate association p-value<0.20 were included in a multivariate logistic model and assessed by forward stepwise multivariate logistic regression analysis.

The algorithm resulting from the multivariate logistic regression was reduced by excluding predictors with p-value >0.10.

The probability of patients having a diagnosis of depression was calculated using the formula 1/(1þe_x) in which x is the sum of the constant and product terms of the regression coefficients and variables representing different covariates. The logistic

regression produced predicted values ranging from 0 to 1, obtained by multiplying the observed values for each independent variable by the coefficients obtained in the regression model. Different possible cut-off probabilities between 0 and 1 were selected to maximize sensitivity and specificity. The amount of patients correctly and incorrectly classified as having a diagnosis of depression for the different cut-off probabilities was determined. Sensitivity (i.e. proportion of people with recorded depression diagnosis correctly classified as such), specificity (i.e. proportion of people without a recorded depression diagnosis correctly classified as such) and positive predictive value (PPV, i.e. posterior probability of having a recorded depression diagnosis given that patients are classified as such according to the algorithm) were computed.

Validation of the diagnostic algorithm

To validate the algorithm internally we investigated its reliability and discrimination.29 The reliability (goodness of fit) refers to the correspondence between estimated probability and observed frequencies which was evaluated by using the Hosmer and Lemeshow test.30 The discrimination, or the ability of our algorithm to separate antidepressant drug users with and without diagnosis of depression, was measured using the area under the receiver operating characteristic (ROC) curve. The area under the ROC curve is the probability that the results are correctly classified, given one antidepressant user with a diagnosis of depression and one without. AUC ranges from 0.5 (no apparent accuracy) to 1.0 (perfect accuracy). The AUC was calculated through the trapezoidal rule.31 For external validation, we assessed the generalizability and applicability of the derived algorithm by applying it to the validation set. The discrimination ability was estimated using ROC curves.

[TABLE 1] [TABLE 2]

RESULTS

Derivation of the algorithm

The demographic and clinical characteristics of the 1855 antidepressant drug users are presented in Table 1. The derivation set was about 67% female with a mean age (standard deviation; SD) of 50 (17) years. More than half of the antidepressant drug users received an SSRI and most prescriptions were prescribed by a GP (84%). About 22% received only one antidepressant prescription during the 12 months following index prescription while 51% received five or more prescriptions during this period. With regards to co-medication 29% of the antidepressant drug users received anxiolytics (ATC N05B), 28% anti-inflammatory agents (ATC M01), 20% analgesics (ATC N02) and 20% hypnotics/sedatives (ATC N05C) in the 6 months prior to index date. Of the 1855 antidepressant drug users, 961 (51.8%) patients had a recorded diagnosis of depression in their medical files.

The independent contributions of different covariates with a p-value below 0.20 are presented in Table 2.

In Table 2, we also present an overview of use of nervous system drugs (ATC group N) in our derivation set. Although age and gender did not show a p-value below 0.20 they were included in the model. The multiple logistic regression showed that age, SSRIs rather than other antidepressants prescribed on the index date, the dose of the second prescription, GP as prescriber of antidepressants and the number of

antidepressant prescriptions dispensed in the 12 months following the index date plus medication dispensed for treating acid related disorders, laxatives, cardiac therapy or hypnotics/sedatives prescribed in the 6 months prior to index date are significant correlates with depression recorded in the GP medical file. In a population of first time antidepressant drug users the probability that an antidepressant drug is prescribed in a patient with recorded depression is therefore: 1/(1þe_x) in which: x¼_4:344þ0:014 _ ageþ0:315_ SSRIþ1:214_ dose^{0}:^{5}_^{0}:^{99DDD}þ1:827_dose_^{1}:^{0} ^{DDD}þ0:489_switchþ 0:881_nr^{2}_^{4} of Rx0s during12 monthsþ1:016_ nr_^{5} of Rx0s during 12 months þ 1:709 _ GP prescriber on index date _ 0:313 _ A02 _ 0:323 _ A06 _ 1:0_ C01 þ 0:363 _ N05C The model coefficients are presented in Table 3. The strongest predictors were GP as prescriber on index date (OR¼5.5, 95%CI 4.0–7.6), higher doses compared to lower doses (OR¼6.2, 95%CI 4.3–9.1 for 1 or more DDDs versus <0.5 DDDs) and receiving more than one antidepressant prescription in 12 months following the first prescription (OR¼2.8, 95%CI 2.1– 3.7 when receiving five or more antidepressant prescriptions).

[TABLE 3] [FIGURE 1] [TABLE 4]

The Hosmer and Leweshow goodness-of-fit test supported the reliability of the algorithm (p¼0.40).

The area under the ROC curve was 0.79 (95%CI, 0.77– 0.81) and is presented in Figure 1. The probability that the algorithm correctly identifies an antidepressant drug user receiving antidepressant as having a recorded diagnosis of depression was 79%. With a cut-off level of 0.5 for the predicted probability of having a recorded diagnosis of depression when receiving an antidepressant prescription the algorithm had a sensitivity of 79.6% and a specificity of 66.9%. The sensitivity, specificity and PPV for the different cut off values are presented in Table 4.

External validation of the algorithm

Table 1 shows the demographic and clinical characteristics of the validation study population. There were several differences between the derivation and validation sets. The antidepressant drug users in the validation set were slightly older and almost all received a prescription from a GP (99.8%). A larger proportion of the antidepressant drug users in the validation set received other antidepressants than SSRI or TCA on the index date and more antidepressant drug users receiving only one antidepressant prescription during the 12 months from index date. Switching of antidepressant drug was also less common than in the derivation set. Of 3231 antidepressant drug users, 1494 (46.2%) had a recorded depression diagnosis in their medical files.

When applied to the validation set population, the depression algorithm showed a discrimination property of 73.0% (95%CI 71.3–74.7%). The area under the ROC curve is presented in Figure 2. The sensitivity and specificity, for a cut-off level of 0.5 for the predicted probability, were 81.5% and 53.0%, respectively.

DISCUSSION

We have developed an algorithm that can be used on prescription data to identify patients on antidepressant drug therapy with a diagnosis of depression. To our best knowledge this is the first depression identifying algorithm that uses prescription data only.

Prescription databases are useful tools to investigate therapeutic treatment and clinical outcomes but often lack information on which diseases or symptoms the drug is being prescribed for. Selecting medicines as a proxy for disease can be problematic as most drugs can be prescribed for more than one illness or symptom. There are differences in the sensitivity of a certain drug for identifying a specific disease.

[FIGURE 2]

Kolodner et al.[21] showed that using a prescription of antimigraine preparations as a marker for migraine has very low sensitivity of 11.1%, while Shackleton et al.[19] reported that using epilepsy medication polytherapy as a marker for epileptic patients shows high sensitivity (79%). As the antidepressants are being prescribed for various illnesses other than depression, the depression algorithm presented here can serve as a useful tool in observational studies that investigate antidepressant treatment outcomes in depressed patients.

The developed algorithm has been shown to exert a discrimination property of 73% and 79% in two populations. Although there is no measurement reference stating whether discrimination ability is 'good enough', a rough guide classifies algorithms with a discrimination ability of 70–80% as fair. In addition, the algorithm is easy to use as it includes only basic variables that can be found in most prescription databases. Applying the algorithm would create a more homogenous study cohort from prescription databases when there is need to study patients suffering from depression. It can serve as a tool for researchers that undertake studies in prescription databases that focus on investigating long-term treatment outcomes for depressed patients.

There are some limitations to our algorithm. Firstly, the algorithm is developed in a population of antidepressant drug users and as a result it cannot identify patients with a depression diagnosis who are not treated with antidepressants. Secondly, although we did validate our algorithm externally in another population in The Netherlands, it might not be valid for every country. In The Netherlands, the GP is a gatekeeper to secondary care. This situation will vary in other countries, where the structure of primary and secondary care differs from that of The Netherlands.

Thirdly, the algorithm is developed and validated in data from 2001 and as treatment guidelines and indication scope of antidepressant drug use broadens throughout the years applying it on older data specifically during the years when the SSRIs were newly marketed is not advisable.With regards to newer data, the validity of the algorithm needs to be ascertained. Fourthly, our algorithm identifies patients on antidepressants that have a depression diagnosis recorded in their medical files. This diagnosis is often made by a GP and may not be consistent with DSM-IV criteria. In this study, we chose a non-strict definition of depression including both patients with codes for symptoms of depression (P03) and the formal diagnosis (P76). Symptoms of depression are frequently a precursor of a depression diagnosis. Only 17% of the antidepressant drug users had registered symptoms of depression (P03) without a subsequent depression diagnosis. In an earlier study, we found that GPs often fail to register the reason for prescribing the antidepressant.[18] We expect a part of these patients to receive a depression diagnosis later on. In addition, the GPs lack of registering the indication for prescribing can lead to a misclassification error that would probably lower the sensitivity of our algorithm. Lastly, the algorithm is based

on incident and not on prevalent users, which limits its use in studies of cross sectional design.

Although we believe that the algorithm can be a useful tool to identify patients with a recorded depression diagnosis there are still gaps that need to be filled when studying depression in prescription databases. A marker for depression severity is still missing. In addition, using antidepressants as a proxy to identify depressed patients leaves out patients suffering from depression that receive other kind of treatment such as psychotherapy.

In conclusion, we have managed to develop and validate an algorithm that can be used to identify depressed patients on antidepressant therapy. The algorithm is a useful tool that can be used to compose cohorts of patients treated for depression from prescription databases.

KEYPOINTS

As antidepressants are being prescribed for various illnesses other than depression, the developed depression algorithm can serve as a useful tool in observational studies that investigate antidepressant treatment outcomes in depressed patients.

_ The predictive algorithm included age, SSRI prescribed on the index date, prescribed dose, GP as prescriber and the number of antidepressant prescriptions prescribed plus medication for treating acid related disorders, laxatives, cardiac therapy or hypnotics/sedatives prescribed in the 6 months prior to index date.

_ The developed algorithm has been shown to exert a discrimination property of 73% and 79% in two populations.

REFERENCES

1. Anderson IM. SSRIS versus tricyclic antidepressants in depressed inpatients: a meta-analysis of efficacy and tolerability. Depress Anxiety 1998; 7 (Suppl 1): 11–17.
2. Rosholm JU, Gram LF, Isacsson G, Hallas J, Bergman U. Changes in the pattern of antidepressant use upon the introduction of the new antidepressants: a prescription database study. Eur J Clin Pharmacol 1997; 52(3): 205–209.
3. Meijer WE, Heerdink ER, Leufkens HG, Herings RM, Egberts AC, NolenWA. Incidence and determinants of long-term use of antidepressants.
Eur J Clin Pharmacol 2004; 60(1): 57–61.
4. Munoz-Arroyo R, Sutton M, Morrison J. Exploring potential explanations for the increase in antidepressant prescribing in Scotland using secondary analyses of routine data. Br J Gen Pract 2006; 56(527): 423– 428.
5. Lawrenson RA, Tyrer F, Newson RB, Farmer RD. The treatment of depression in UK general practice: selective serotonin reuptake inhibitors and tricyclic antidepressants compared. J Affect Disord 2000; 59(2): 149–157.
6. Fairman KA, Drevets WC, Kreisman JJ, Teitelbaum F. Course of antidepressant treatment drug type, and prescriber's specialty.
Psychiatr Serv 1998; 49(9): 1180–1186.
7. Olfson M, Marcus SC, Pincus HA, Zito JM, Thompson JW, Zarin DA.
Antidepressant prescribing practices of outpatient psychiatrists. Arch Gen Psychiatry 1998; 55(4): 310–316.
8. Percudani M, Barbui C, Fortino I, Petrovich L. Antidepressant drug prescribing among elderly subjects: a population-based study. Int J Geriatr Psychiatry 2005; 20(2): 113–118.
9. van Eijk ME, Bahri P, Dekker G, et al. Use of prevalence and incidence measures to describe age-related prescribing of antidepressants with and without anticholinergic effects. J Clin Epidemiol 2000; 53(6): 645– 651.
10. Hemels ME, Koren G, Einarson TR. Increased use of antidepressants in Canada 1981–2000. Ann Pharmacother 2002; 36(9): 1375–1379.

11. Egberts ACG, Stuijt CCM, Heerdink ER, Leufkens HGM. Indicaties voor het begruik van antidepressiva [Indications for antidepressant drug use]. Pharm Weekblad 1998; 133(20): 776–780.
12. Walsh JK. Pharmacologic management of insomnia. J Clin Psychiatry 2004; 65 (Suppl 16): 41–45.
13. Zinner NR, Koke SC, Viktrup L. Pharmacotherapy for stress urinary incontinence: present and future options. Drugs 2004; 64(14): 1503– 1516.
14. Colombo B, Annovazzi PO, Comi G. Therapy of primary headaches: the role of antidepressants. Neurol Sci 2004; 25 (Suppl 3): S171–S175.
15. Farmacotherapeutisch Kompas 2005. De Commissie Farmaceutische Hulp van het College voor zorgverzekeringen. CVZ: Amstelveen, 2005.
16. Maizels M, McCarberg B. Antidepressants and antiepileptic drugs for chronic non-cancer pain. Am Fam Physician 2005; 71(3): 483–490.
17. Saad RJ, Chey WD. Review article: current and emerging therapies for functional dyspepsia. Aliment Pharmacol Ther 2006; 24(3): 475–492.
18. Gardarsdottir H, Heerdink ER, van Dijk L, Egberts AC. Indications for antidepressant drug prescribing in general practice in The Netherlands.
J Affect Disord 2007; 98(1–2): 109–115.
19. Shackleton DP, Westendorp RG, Kasteleijn-Nolst Trenite DG, de Boer A, Herings RM. Dispensing epilepsy medication: a method of determining the frequency of symptomatic individuals with seizures. J Clin Epidemiol 1997; 50(9): 1061–1068.
20. van de Vijver DA, Stricker BH, Breteler MM, Roos RA, Porsius AJ, de BoerA. Evaluation of antiparkinsonian drugs in pharmacy records as a marker for Parkinson's disease. Pharm World Sci 2001; 23(4): 148–152.
21. Kolodner K, Lipton RB, Lafata JE, et al. Pharmacy and medical claims data identified migraine sufferers with high specificity but modest sensitivity. J Clin Epidemiol 2004; 57(9): 962–972.
22. Moth G, Vedsted P, Schiotz P. Identification of asthmatic children using prescription data and diagnosis. Eur J Clin Pharmacol 2007; 63(6): 605–611.
23. Spettell CM,Wall TC, Allison J, et al. Identifying physician-recognized depression from administrative data: consequences for quality measurement. Health Serv Res 2003; 38(4): 1081–1102.
24. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health 2005; 15(1): 59–65.
25. Lamberts H, Wood M. International Classification of Primary Care.
Oxford University Press: Oxford, 1987.
26. Florentinus SR, Souverein PC, Griens FA, Groenewegen PP, Leufkens HG, Heerdink ER. Linking community pharmacy dispensing data to prescribing data of general practitioners. BMC Med Inform Decis Mak 2006; 6: 18.
27. Anatomical Therapeutic Chemical (ATC) Classification Index. WHO Collaborating Centre for Drug Statistics Methodology: Oslo, 2002.
28. Buurma H, Bouvy ML, De Smet PA, Floor-Schreudering A, Leufkens HG, Egberts AC. Prevalence and determinants of pharmacy shopping behaviour. J Clin Pharm Ther 2008; 33(1): 17–23.
29. Harrell FE, Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15(4): 361–387.
30. Hosmer DW, Lemeshow S. Applied Logistic Regression. JohnWiley & Sons: New York, 1989.
31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143(1): 29–36.

TABLES AND FIGURES

Table 1. Basic demographics of the study population, derivation and validation set

	Derivation set ($n = 1855$)	Validation set ($n = 3231$)	p-Value
Patient characteristics			
Female gender, n (%)	1235 (66.6)	2112 (65.4)	0.86
Age (years), mean (SD)	50.1 (16.8)	49.1 (16.4)	0.52
Antidepressant (AD) characteristics			
SSRI, n (%)	1052 (56.7)	1678 (51.9)	
TCA, n (%)	550 (29.7)	790 (24.5)	
Other*, n (%)	253 (13.6)	763 (23.6)	<0.01
Number of AD prescriptions in 1 year			
Only one AD prescription	414 (22.3)	1067 (33.0)	
Two to four AD prescriptions	490 (26.4)	1014 (31.4)	
Five or more AD prescriptions	951 (51.3)	1150 (35.6)	<0.01
Second AD is different from first AD (switch)	53 (5.9)	69 (2.1)	<0.01
Dose			
<0.5 DDDs per day, n (%)	426 (23.0)	687 (21.3)	
0.5–0.99 DDDs per day, n (%)	218 (11.7)	620 (19.2)	
1.0 DDD or higher per day, n (%)	1211 (65.3)	1924 (59.5)	<0.01
Type of prescriber (index date)			
General practitioner	1555 (83.8)	3224 (99.8)	
Specialist/other/unknown	300 (16.2)	7 (0.2)	<0.01

*Other antidepressants: moclobemide, mianserin, trazodone, mirtazapine, venlafaxine.

Table 2. Antidepressant users ($n = 1855$) from the derivation set divided into those with and without a recorded diagnosis of depression

	Recorded depression ($n = 961$)	No recorded depression ($n = 894$)	Statistics (p)*
Patient characteristics			
Female gender, *n* (%)	638 (66.4)	597 (66.8)	0.859
Age (years), mean (SD)	50.3 (17.0)	49.7 (16.6)	0.519
Antidepressant (AD) characteristics			
SSRI, *n* (%)	694 (72.2)	358 (40.0)	
TCA, *n* (%)	131 (13.6)	419 (46.9)	
Other†, *n* (%)	136 (14.2)	117 (13.1)	<0.001
Number of AD prescriptions dispensed in 1 year			
Only one AD prescription	131 (13.6)	283 (31.7)	
Two to four AD prescriptions	249 (25.9)	241 (26.9)	
Five or more AD prescriptions	581 (60.5)	370 (41.4)	<0.001
Second AD is different from first AD (switch)	53 (5.9)	24 (2.7)	0.002
Dose			
<0.5 DDDs per day	80 (8.3)	346 (38.7)	
0.5–0.99 DDDs per day	93 (9.7)	125 (14.0)	
1.0 DDD or higher per day	788 (82.0)	423 (47.3)	<0.001
Type of prescriber (index date)			
General practitioner	896 (93.2)	659 (73.7)	
Specialist/other/unknown	65 (6.8)	235 (26.3)	<0.001
Medication use 6 months prior to index date (ATC Group)			
Acid related disorders (A02), *n* (%)	154 (16.0)	180 (20.1)	0.021
Laxatives (A06), *n* (%)	80 (8.3)	101 (11.3)	0.031
Diabetic drugs (A10), *n* (%)	45 (4.7)	56 (6.3)	0.134
Cardiac therapy (C01), *n* (%)	33 (3.4)	55 (6.2)	0.006
Anti-inflammatory and antirheumatic agents (M01), *n* (%)	230 (23.9)	292 (32.7)	<0.001
Anaesthetics (N01), *n* (%)	7 (0.7)	12 (1.3)	0.189
Analgesics (N02), *n* (%)	168 (17.5)	213 (23.8)	0.001
Antiepileptics (N03), *n* (%)	20 (2.1)	33 (3.7)	0.038
Parkinson medicines (N04), *n* (%)	5 (0.5)	9 (1.0)	0.226
Antipsychotics (N05A), *n* (%)	34 (3.5)	30 (3.4)	0.830
Anxiolytica (N05B), *n* (%)	296 (30.8)	246 (27.5)	0.120
Hypnotics/sedatives (N05C), *n* (%)	211 (22.0)	159 (17.8)	0.025
Other nervous system drugs (N07), *n* (%)	42 (4.4)	31 (3.5)	0.318
Medication used 6 months past index date (ATC group)			
Acid related disorders (A02)	158 (16.4)	174 (19.5)	0.090
Laxatives (A06)	95 (9.9)	110 (12.3)	0.097
Diabetic drugs (A10), *n* (%)	47 (4.9)	56 (6.3)	0.197
Cardiac therapy (C01), *n* (%)	32 (3.3)	45 (5.0)	0.066
Anti-inflammatory and antirheumatic agents (M01), *n* (%)	199 (20.7)	258 (28.9)	<0.001
Anaesthetics (N01), *n* (%)	10 (1.0)	13 (1.5)	0.421
Analgesics (N02), *n* (%)	181 (18.8)	200 (22.4)	0.060
Antiepileptics (N03), *n* (%)	25 (2.6)	48 (5.4)	0.002
Parkinson medicines (N04), *n* (%)	7 (0.7)	9 (1.0)	0.517
Antipsychotics (N05A), *n* (%)	49 (5.1)	44 (4.9)	0.861
Anxiolytica (N05B), *n* (%)	290 (30.2)	260 (29.1)	0.606
Hypnotics/sedatives (N05C), *n* (%)	207 (21.5)	162 (18.1)	0.065
Other nervous system drugs (N07), *n* (%)	26 (2.7)	27 (3.0)	0.684

The *p*-values are originated from univariate analysis of each covariate.
*Chi square statistics.
†Other antidepressants: moclobemide, mianserin, trazodone, mirtazapine, venlafaxine.

Table 3. Variables from multivariate logistic regression of the derivation set that predict a registered depression diagnosis in the antidepressant users medical file

Variables	β	SE	Exp (β)	95%CI
Patient characteristics				
Age	0.014	0.004	1.014	1.007–1.021
Antidepressant (AD) characteristics				
SSRI	0.315	0.145	1.370	1.014–1.852
Number of DDDs				
<0.5 DDDs per day			Ref	
0.5–0.99 DDDs per day	1.214	0.199	3.367	2.279–4.975
1.0 DDD or higher per day	1.827	0.192	6.216	4.270–9.048
Second AD is different from first AD (switch)	0.489	0.274	1.631	0.953–2.793
Number of AD prescriptions dispensed in 1 year				
Only one AD prescription			Ref	
Two to four AD prescriptions	0.881	0.159	2.412	1.767–3.293
Five or more Ad prescriptions	1.016	0.143	2.763	2.089–3.655
Type of prescriber (index date)				
AD prescribed by a GP	1.709	0.163	5.524	4.013–7.605
Medication use 6 months prior to index date				
Acid related disorders (A02)	−0.313	0.146	0.731	0.550–0.973
Laxatives (A06)	−0.323	0.191	0.724	0.498–1.052
Cardiac therapy (C01)	−1.000	0.268	0.386	0.217–0.622
Hypnotics and Sedatives (N05C)	0.363	0.141	1.438	1.090–1.896
Constant	−4.344	0.303	0.013	

SE, standard error.

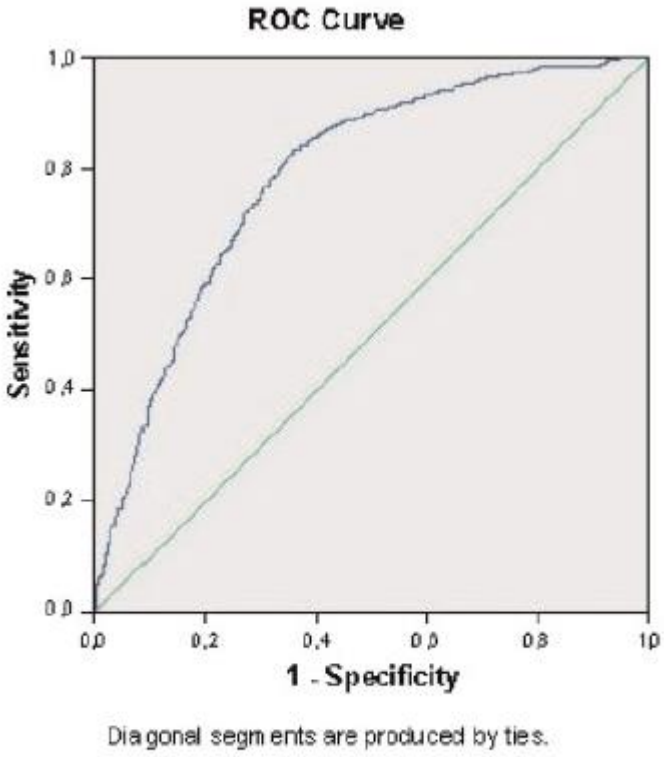

Figure 1. ROC curve for the multivariate logistic regression model predicting diagnosis of depression in a population of first time antidepressant users. The curve shows sensitivity versus 1-specificity based on probabilities computed through multivariable logistic regression. Area under curve = 0.79

Table 4. Sensitivity, specificity and positive predictive value of the algorithm

Probability	Recorded depression	No recorded depression	Sensitivity (%)	Specificity (%)	PPV (%)
≥0.1	949	773	98.7	13.5	55.1
≥0.2	930	646	96.8	27.7	59.0
≥0.3	885	500	92.1	44.1	63.9
≥0.4	832	365	86.6	59.1	69.5
≥0.5	765	296	79.6	66.9	72.1
≥0.6	671	236	69.8	73.6	74.0
≥0.7	531	161	55.2	82.0	76.7
≥0.8	137	26	14.3	97.1	84.0

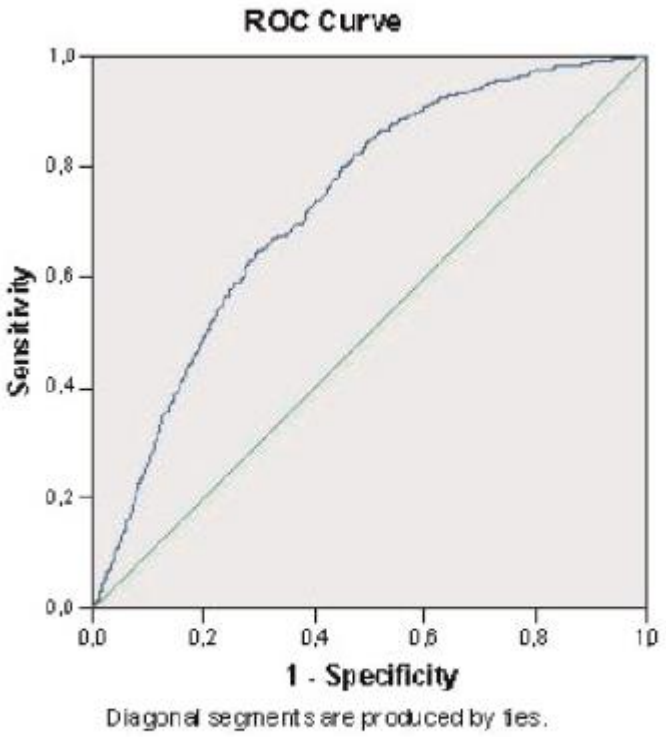

Figure 2. ROC curve for the validation set predicting diagnosis of depression in a population of first time antidepressant users. The curve shows sensitivity versus 1-specificity based on probabilities computed through multivariable logistic regression. Area under curve = 0.73