Chapter 3

Do GPs' medical records demonstrate a good recognition of depression? A new perspective on case extraction

Karlijn J Joling
Harm WJ van Marwijk
Ellen Piek
Henriëtte E van der Horst
Brenda W Penninx
Peter Verhaak
Hein PJ van Hout

ABSTRACT

Background: Previous estimates of depression recognition in primary care are low and inconsistent. This may be due to registration artifacts and limited extraction efforts. This study investigated a) whether GPs' medical records demonstrate an accurate recognition of depression and b) which combinations of indications within the record most accurately reflect a diagnosis of depression.

Methods: GPs' registrations were compared with a reference standard, the Composite International Diagnostic Interview (CIDI), according to DSM-IV criteria. Six definitions of GPs' recognition of depression were tested using diagnostic codes, medication data, referral data and free text in the medical records. The Youden-index was used to select the optimal definition of recognition. Data were derived from the Netherlands Study of Depression and Anxiety. 816 primary care patients from 33 general practitioners were included in the vicinities of Amsterdam and Leiden, The Netherlands.

Results: Registration of antidepressant prescriptions was the best single indicator of GPs' recognition of CIDI depression with a recognition rate of 0.43. The best combination of indicators increased the recognition rate to 0.69. All indications except the specific diagnostic codes for 'depressive disorder' and 'depressive feelings' were included in this definition.

Limitations: Potential bias due to the selection of participating GPs might have influenced our recognition rates.

Conclusion: GPs are aware of mental health problems in most depressed patients, but labeling with specific diagnostic codes is weak. Researchers should consider that diagnostic coding alone is not an accurate measure of the diagnostic ability of depression and strongly underestimates the accuracy of the GP.

INTRODUCTION

Most care for depressed patients is delivered in general practice.[1-5] Research shows that General Practitioners (GPs) fail to diagnose a large number of cases of mental illness. Recognition rates vary greatly between individual studies, but overall GPs correctly identified depression in less than 50%.[6-12] If these estimates are correct, a considerable proportion of depressed patients are overlooked, particularly in the short term.[13, 14]

However, the low recognition and registration rates of GPs reported in previous studies are perhaps underestimates. Studies examining medical records mostly based their estimates on a registration code as evidence of the GPs' diagnosis.[15, 16] These estimates may not reflect the actual diagnostic process in primary care. Depressed patients might be labeled by their GP as having other psychiatric problems, or the registration code could have been limited to a symptom code ('feeling depressed'), to psychosocial problems or to a prescription of an antidepressant.[17, 18] Other previous studies have used a simple questionnaire or form to assess whether GPs are able to diagnose depression.[17, 19-21] Diagnostic sensitivity in these studies was slightly greater than in the medical-record studies [10], but recognition rates could have artificially been raised through cueing. Furthermore, these results may not reflect 'real' clinical diagnoses, as the registration sheets were completed for research purposes.

It is unclear to what extent the recognition by GPs is better reflected when all available medical record data are taken into account. The aim of this study was to answer the following research questions: 1) How accurately do GPs' records of indications of depression or mental health problems reflect diagnoses of depression? and 2) Which combinations of indications within the medical record most accurately reflect a diagnosis of depression?

METHODS

Study design, procedures and participants

This study compared diagnoses of depression by GPs during routine care contacts with a reference standard to diagnose depressive disorders. GPs, patients and interviewers were blinded to the diagnosis by the reference standard.

Data were derived from the baseline measurements of the Netherlands Study of Depression and Anxiety (NESDA), a large longitudinal cohort study among 2981 participants aged 18 to 65 years. NESDA has been described in detail elsewhere.[22]

The study protocol was approved centrally by the Ethical Review Board of the VU University Medical Centre and subsequently by local review boards for each participating center. After full verbal and written information about the study, written informed consent was obtained from all participants at the start of baseline assessment. Recruitment for NESDA took place between September 2004 and February 2007 in the general population, in general practices and in mental health organizations. We limit our analysis to the respondents recruited from the general practices. A three-stage screening procedure was used for the selection of respondents. Screening questionnaires were sent to a random sample of 23,750 patients who consecutively consulted their GP in the last four months irrespective of the reason for the consultation. Of the screeners returned, the screen-positives were approached for a telephone screen interview consisting of the Composite Interview Diagnostic Interview short form (CIDI-SF).[23]

Those who met the criteria for a current anxiety or depressive disorder, who were not being treated for psychiatric conditions in a psychiatric mental health care setting and provided informed consent were included in the NESDA study (n=898). In addition, a random selection of the screen-negatives was also invited to participate. This resulted in a sample of 1610 participants.[22] For the present analyses, data on 794 patients were not examined, because 1) the patient refused the use of their Electronic Medical Record (EMR, n=207) or 2) there were no free-text data available from the GP (n=587). Free-text is part of the EMR and is used for notes about GP and patient encounters. The availability of the free-text data depended on approval from general practices or accredited Medical Ethical Committees (MECs). The MEC in Groningen – one of the three study locations – did not approve the use of the free-text data. As a result, data from 816 participants were analyzed in this study.

Measurements

Reference standard

DSM-IV diagnoses obtained from the Composite Interview Diagnostic Interview (CIDI) were considered as the reference standard. Patients were regarded as depressed in the event of a diagnosis of current MDD and/or current dysthymia (i.e. in the past year). The CIDI is used worldwide and WHO field research has found high inter-rater reliability, high test–retest reliability and high validity for depressive and anxiety disorders.[24-26] Symptoms of depression and their severity were also assessed using the Inventory of Depressive Symptomatology-Self Report (IDS-SR30).[27]

Diagnosis of depression by the GPs

Data from the GP diagnoses were extracted from the EMRs starting from one year prior to inclusion in NESDA to one year after inclusion. Different sets of registration characteristics were explored to estimate the best accuracy of the GPs' detection of depressive disorders. The following characteristics could be part of the various definitions:

Registered diagnoses: Diagnoses were recorded by GPs according to the International Classification of Primary Care (ICPC) codes.[28] GPs were trained in ICPC coding and met regularly to support uniformity in coding. The ICPC codes 'depressive disorder (P76)' and 'depressive feelings (P03)' were used as registered diagnoses of depression. The diagnostic criteria for depressive disorder are based on the DSM-IV criteria for major depression. Furthermore, we made a selection of other ICPC codes within the ICPC-Psychological Chapter: "Other psychological-codes" (Appendix A).

Antidepressant prescriptions: Prescriptions for medication were coded by GPs according to the Anatomical Therapeutic Chemical classification system (ATC). Antidepressant medication included all ATC codes listed under N06A.

Referrals to mental health care: Referrals to mental health care were registered by the GPs with use of Working Committee for Information and Automation (WCIA) codes. We included referrals to (primary care) psychologists, social work, psychiatry, psychotherapy and to a mental health care institution.

Free-text indications of depression: Part of the EMR is available for the GP to record information as free text. The free text of consultations was scored for notes on depression and other mental health and psycho-social problems. All phrases about mental health and psychosocial problems were listed from an initial analysis of the free-text of 30 medical records. On the basis of this list and according to the diagnostic criteria of the DSM-IV, a scoring system was created to score free-text data consistently. Indications for depression involved keywords relating to the two main symptoms of major depression according to the DSM-IV. Inter-rater reliability between two scorers was measured for 36 cases and there was total agreement (kappa 1.00), after which all free-text parts were scored by one of the scorers.

Statistical analysis

We compared the reference standard with the following six characteristics registered by the GPs 1) registration of P76, 2) registration of P03, 3) registration of other psychological codes, 4) antidepressant prescription, 5) referral to mental health care or 6) a free-text indication for depression. Sensitivity and specificity of each of the characteristics were calculated. The Youden index (Y, sensitivity + specificity −1) was used to determine the most accurate definition of recognition.[29]

To determine the most accurate definition of recognition we first selected the registration characteristic with the highest Y and added each of the other six characteristics separately to this variable. The combination of characteristics with the highest Y was selected and again each of the other remaining characteristics was added. This process was repeated until Y was at its maximum. This combination of registration characteristics was defined as the most accurate definition of recognition.

Because the study sample was composed of three subgroups weighting factors were used to recalculate the study population back to the baseline population. The calculation of the weighting factors has been described in detail elsewhere.[30]

RESULTS

Characteristics of the study sample and GPs

The 816 participants were recruited by 33 GPs from 12 general practices in the vicinities of Amsterdam and Leiden, The Netherlands. The mean age of the participants was 46.17 years (SD 12.3, range 18–65) and 68.0% were female (68.0%). Their educational level differed between basic (6.6%), intermediate (55.9%) and higher education (37.5%). Participants were less likely to have attended higher education (χ^2=9.89, df 2, p=0.007) and had less often experienced lifetime depression (not current) according to the CIDI (χ^2=4.11, df 1, p=0.043) than patients who were excluded because of the lack of free-text data. The participants did not differ by age, sex and current depression CIDI diagnoses from the excluded patients.

Baseline characteristics from 31 of the 33 GPs participating in the study were available. Their mean age was 48.5 (SD 8.3) and thirteen were female (41.9%). They had on average 17.8 (SD 9.6) years experience and fourteen (45.2%) of them had a special interest in depression. The number of analyzed patients per GP ranged from 1 to 51 (mean 24.7, SD 13.3).

Recognition of depression by GPs

According to the reference standard, the CIDI, 263 of 816 participants (32.2%) were depressed: 196 patients had MDD alone, 15 had dysthymia alone and 52 had both MDD and dysthymia. Comorbidity of depression and anxiety was present in 179 patients. Of the 553 patients without current depression, 180 persons had experienced MDD or dysthymia in their lifetime. When comparing the six characteristics separately with CIDI-depression diagnoses, the registration of antidepressant prescriptions yielded the highest accuracy with a recognition rate of 0.43 (Table 1).

Based on the registration of antidepressant prescriptions, 107 of the 263 depressed patients were recognized as depressed by GPs, while 156 depressed patients were not recognized.

Table 1: Sensitivity, specificity and Youden index (Y) of the 6 indications of depression

Indication of depression	N	Se	Sp	Y
Antidepressant prescription	164	0.433	0.944	0.377
Strong free-text indication of depression	134	0.407	0.956	0.363
Referral to mental health care	156	0.368	0.932	0.301
ICPC-code P76 "Depressive disorder"	83	0.258	0.983	0.241
ICPC Psychological-codes (other than P03 and P76)	134	0.201	0.905	0.107
ICPC-code P03 "Depressive feelings"	28	0.070	0.983	0.053

Abbreviations: Se, sensitivity; Sp, specificity; Y, Youden index; ICPC, International Classification of Primary Care; P-codes, ICPC-codes from the Psychological chapter including codes for psychological problems and mental disorders

We examined the benefit of adding characteristics to this definition of recognition in order to determine the combination of characteristics with the best accuracy. This definition maximized the recognition rate to 0.69 with a specificity of 0.81 and included the registration of antidepressant prescriptions, significant free-text indications, referrals to mental health care and other psychological-codes (Table 2). Based on this definition, 182 of the 263 depressed patients were recognized as depressed by GPs, while 81 depressed patients were not recognized. Of the 553 without current depression, 157 patients were diagnosed by their GP as depressed according to this definition. Of these 157 false positives, 65 (41.4%) were falsely detected, because they were coded under depression with ICPC code P76 (n=8) or an antidepressant prescription (n=49) or both characteristics (n=16). Adding the registration of ICPC codes P76 or P03 to the definition did not result in better accuracy (Table 2).

Table 2: The effect of adding indications of depression with the highest diagnostic agreement to recognition accuracy

Definition of recognition	Se	Sp	Y
1. Antidepressant prescription	0.433	0.944	0.377
2. Definition 1 + strong free-text indication of depression	0.564	0.914	0.478
3. Definition 2 + referral to mental health care	0.638	0.859	0.497
4. Definition 3 + ICPC Psychological-codes (other than P03 and P76)	0.693	0.811	0.505
5. Definition 4 + ICPC P76 "Depressive disorder"	0.695	0.808	0.503
6. Definition 5 + ICPC P03 "Depressive feelings"	0.698	0.804	0.502

Abbreviations: Se, sensitivity; Sp, specificity; Y, Youden index; ICPC, International Classification of Primary Care; P-codes, ICPC-codes from the Psychological chapter including codes for psychological problems and mental disorders

Recognition of severe cases

Recognition improved for patients who were more severely depressed. Among the patients given a CIDI diagnosis with moderate to very severe depression (IDS-SR30 scores ≥26) the recognition rate was 75% according to the best definition from Table 2 and 45.9% according to the best single indicator of recognition. GPs' recognition of severe or very severely depressed patients (IDS-SR30 scores ≥39) increased respectively to 80.4% and to 57.1%.

Variation between GPs

We examined the variation in recognition rates between general practices based on the best definition of recognition. Sensitivity ranged between 0.40 and 1.00 and specificity ranged between 0.14 and 0.84. The GPs with a special interest in depression recognized 71.1% of the cases according to the best definition while the other GPs recognized 68.3%. When using the best single indicator as criterion the recognition rate for GPs with a special interest was 0.46 versus 0.38 for the other GPs.

DISCUSSION

Summary of main findings

One aim of this study was to report how accurately GPs' records of indications of depression reflect the diagnosis of depression and to evaluate which combinations of indications were the most accurate. According to the most optimal definition of recognition, 69.3% of the depressed patients were identified by the GPs, while they could accurately exclude 81.1% of the non-depressed patients. Recognition even increased to 80% for the (very) severely depressed cases. The most accurate definition included the registrations of antidepressants prescriptions, a significant indication of depression in the free-text, referrals to mental health care, and registration of relevant ICPC-codes for psychological problems and mental disorders other than the specific codes for depressive disorders (P76) and depressive feelings (P03). Although several of these indicators do not reflect a real diagnosis of depression it indicates the extent that the records reflect an overall concern the GP had and acted on about mental illnesses.

Of all the separate indications, registrations of antidepressants prescriptions resulted in the highest accuracy of recognition. The diagnostic codes for depressive disorders (P76), depressive feelings (P03) and other psychological problems (P-codes) added least to the most optimal definition of recognition.

While most GPs use these codes, all other registered characteristics performed better, also as standalone indicators. This indicates that ICPC codes may still not be well documented by GPs or that they may have a lot of "tacit knowledge" that is not explained by the diagnostic codes.

Comparison with existing literature
The recognition rate of 0.69 we found was much higher than reported in previous studies, which have mostly found that less than half of the patients with depression were recognized by their primary care physicians.[6, 8, 10] A large amount of these studies used stricter definitions of recognition and did not take into account the registration of antidepressant prescriptions, referrals or indications for depression in the free-text of the EMR. This may indicate that diagnostic coding by GPs is not an accurate measure of the diagnostic ability of depression and strongly underestimates the accuracy of the GP. Low recognition might be due to a lack of registration rather than to a lack of awareness.

Strengths and limitations of the study
This study had several strengths. We used an innovative method in order to provide a new view on the previously published recognition estimates. Furthermore, we were also able to make use of a large sample of primary care patients. We analyzed the detection of depression when GPs make routine diagnoses and were able to derive several registration characteristics from their EMRs over time. In contrast to most other studies, we also derived documented information on depression and other mental and psycho-social problems from the free-texts of the EMR. The GP registrations were compared with the CIDI, which is assumed to be the state of the art reference standard and GPs, patients and interviewers were also blinded to the CIDI diagnoses.

Besides these strengths, this study also had some limitations. Potential bias due to the selection of participating GPs might have influenced our recognition rates. GPs in this study were willing to participate in research about depression and 45% of them even had 'a special interest in depression', which could mean that their awareness of depression was higher. However, the recognition rate was found to be quite similar in GPs with and without a special interest (0.71 versus 0.68). GPs were perhaps relatively better trained in ICPC coding and supported to uniformity in coding than the average GP in The Netherlands. Therefore, we may have overestimated recognition, but this is not very likely, as the documented ICPC codes added least to the most optimal definition of recognition.

Although the recognition rate we found is higher than reported in other studies, it might still be an underestimation. Patients were regarded as depressed if there had been a CIDI diagnosis in the past year, but we did not analyze when the episode had manifested during this year and whether the patient had really visited the GP during the episode.

Another limitation is related to the time period we extracted data from the EMRs. The GP diagnoses during the period one year after inclusion in NESDA were also included in the calculations of the sensitivity and specificity. This longer time period could have overestimated the number of false negatives and false positives. Patients with a depression CIDI diagnosis could have recovered or remitted during the period after the CIDI interview. Besides, non-depressed patients could still have developed depression and have been recorded as depressed by the GP. However, the longer time period may be seen as a strength as well. Because patients visit their GP irregularly, some time was also allowed for the GP to detect depression. In general practice, a diagnosis is a process rather than an assessment and often occurs over time over several consultations, taking place after rather than before management and treatment decisions.[31, 32] This indicates the need to look at the narrative through the records of how a mental illness diagnosis is constructed, and requires a more detailed study than was undertaken here.

A last limitation includes the fact that we only compared GPs' registrations with the presence of a depression or dysthymia CIDI diagnosis and did not take into account the presence of other possible psychiatric diagnosis. Patients without a CIDI depression, but diagnosed by the GP as having anxiety, for instance, have now been defined as non-cases according to the CIDI. This might have resulted in an underestimation of the specificity.

Conclusion and implications

In conclusion, our findings imply that GPs are more aware of mental health problems in their patients who are depressed than previously reported. Therefore, it could be argued that training around recognition is less of a priority. This study shows that just using the specific diagnostic codes to determine the ability of GPs' recognition of mental health problems of their patients will not provide accurate estimates. Researchers should critically consider how to extract their data from medical records for future research on mental health problems in primary care.

Appendix A

The following ICPC codes from the ICPC-Psychological Chapter were included in the "other psychological problems" indicator:

P01 feeling anxious/nervous/tense/inadequate, P02 acute stress/transient situational disturbance, P04 feeling/ behaving irritable/angry, P06 disturbances of sleep/insomnia, P27 fear of mental disorder, P73 affective psychosis, P74 anxiety disorder/state, P75 hysterical/hypochondriacal disorder, P77 suicide attempt, P78 neurasthenia, surmenage, P79 other neurotic disorder, P82 posttraumatic stress disorder and P86 anorexia nervosa /bulimia.

REFERENCES

1   Bijl RV, Ravelli A. Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. *Am J Public Health* 2000 April;90(4):602-7.

2   Harman JS, Veazie PJ, Lyness JM. Primary care physician office visits for depression by older Americans. *J Gen Intern Med* 2006 September;21(9):926-30.

3   Kessler D, Sharp D, Lewis G. Screening for depression in primary care. *Br J Gen Pract* 2005 September;55(518):659-60.

4   Tiemens BG, Ormel J, Simon GE. Occurrence, recognition, and outcome of psychological disorders in primary care. *Am J Psychiatry* 1996 May;153(5):636-44.

5   Wittchen HU, Hofler M, Meister W. Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *Int Clin Psychopharmacol* 2001 May;16(3):121-35.

6   Cepoiu M, McCusker J, Cole MG, Sewitch M, Belzile E, Ciampi A. Recognition of depression by non-psychiatric physicians--a systematic literature review and meta-analysis. *J Gen Intern Med* 2008 January;23(1):25-36.

7   Hirschfeld RM, Keller MB, Panico S et al. The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *JAMA* 1997 January 22;277(4):333-40.

8   Jackson JL, Passamonti M, Kroenke K. Outcome and impact of mental disorders in primary care at 5 years. *Psychosom Med* 2007 April;69(3):270-6.

9   Klinkman MS, Coyne JC, Gallo S, Schwenk TL. False positives, false negatives, and the validity of the diagnosis of major depression in primary care. *Arch Fam Med* 1998 September;7(5):451-61.

10 Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009 August 22;374(9690):609-19.

11 Rost K, Zhang M, Fortney J, Smith J, Coyne J, Smith GR, Jr. Persistently poor outcomes of undetected major depression in primary care. *Gen Hosp Psychiatry* 1998 January;20(1):12-20.

12 Schmaling KB, Hernandez DV. Detection of depression among low-income Mexican Americans in primary care. *J Health Care Poor Underserved* 2005 November;16(4):780-90.

13 ten Have M., de Graaf R., Vollebergh W, Beekman A. What depressive symptoms are associated with the use of care services? Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J Affect Disord* 2004 June;80(2-3):239-48.

14 Wittchen HU, Pittrow D. Prevalence, recognition and management of depression in primary care in Germany: the Depression 2000 study. *Hum Psychopharmacol* 2002 June;17 Suppl 1:S1-11.

15 Nuyen J, Volkers AC, Verhaak PF, Schellevis FG, Groenewegen PP, van den Bos GA. Accuracy of diagnosing depression in primary care: the impact of chronic somatic and psychiatric co-morbidity. *Psychol Med* 2005 August;35(8):1185-95.

16 Smolders M, Laurant M, Verhaak P et al. Adherence to evidence-based guidelines for depression and anxiety disorders is associated with recording of the diagnosis. *Gen Hosp Psychiatry* 2009 September;31(5):460-9.

17 Tiemens BG, VonKorff M, Lin EH. Diagnosis of depression by primary care physicians versus a structured diagnostic interview. Understanding discordance. *Gen Hosp Psychiatry* 1999 March;21(2):87-96.

18 Volkers A, de Jong A, de Bakker D, van Dijk L. Doelmatig voorschrijven van antidepressiva in de huisarts-praktijk. Utrecht: NIVEL; 2005.

19 Balestrieri M, Baldacci S, Bellomo A et al. Clinical vs. structured interview on anxiety and affective disorders by primary care physicians. understanding diagnostic discordance. *Epidemiol Psichiatr Soc* 2007 April;16(2):144-51.

20 Menchetti M, Belvederi MM, Bertakis K, Bortolotti B, Berardi D. Recognition and treatment of depression in primary care: effect of patients' presentation and frequency of consultation. *J Psychosom Res* 2009 April;66(4):335-41.

21 Ormel J, van den Brink W, Koeter MW et al. Recognition, management and outcome of psychological disorders in primary care: a naturalistic follow-up study. *Psychol Med* 1990 November;20(4):909-23.

22 Penninx BW, Beekman AT, Smit JH et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* 2008;17(3):121-40.

23 Kessler RC, Andrews G, Mroczek D, Ustun TB, Wittchen HU. The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research* 1998;7(4):171-85.

24 Wacker HR, Battegy R, Mullejans R, Schlosser C. Using the CIDI-C in the general population. In: Stefanis CN, Rabavilas AD, Soldatos CR, editors. *Psychiatry: a world perspective.*Amsterdam: Elsevier Science Publishers; 1990. p. 138-43.

25 Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D. Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. *Br J Psychiatry* 1991 November;159:645-53, 658.

26 Wittchen HU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994 January;28(1):57-84.

27 Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Res* 1986 May;18(1):65-87.

28 WONCA. ICPC-2 International Classification of Primary Care. Oxford: Oxford University Press; 1998. Report No.: 2.

29 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3(1):32-5.

30 Donker T, Comijs H, Cuijpers P et al. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res* 2010 March 30;176(1):45-50.

31 Howie JG. Further observations on diagnosis and management of general practice respiratory illness using simulated patient consultations. *Br Med J* 1974 June 8;2(5918):540-3.

32 Kessler D, Bennewith O, Lewis G, Sharp D. Detection of depression and anxiety in primary care: follow up study. *BMJ* 2002 November 2;325(7371):1016-7.