Psychological Medicine, 1999, **29**, 833–845. Printed in the United Kingdom
© 1999 Cambridge University Press

Training primary-care physicians to recognize, diagnose and manage depression: does it improve patient outcomes?

B. G. TIEMENS,[1] J. ORMEL, J. A. JENNER, K. VAN DER MEER, T. W. D. P. VAN OS, R. H. S. VAN DEN BRINK, A. SMIT AND W. VAN DEN BRINK

From the Department of Psychiatry and Department of Family Medicine, University of Groningen; and Department of Psychiatry, University of Amsterdam, The Netherlands

ABSTRACT

Background. We developed a comprehensive, 20-hour training programme for primary-care physicians, that sought to improve their ability to detect, diagnose and manage depression. We evaluated the effects of physician training on patient outcomes, using a pre-post design.

Methods. In the pre-training phase of the study, we sampled 1834 consecutive patients of 17 primary-care physicians and evaluated 518 of these patients for the presence of depression. We measured outcomes of all patients with depression at 3 months and 1 year. The outcome measures were: severity of psychopathology; duration of depressive episode; and level of daily functioning. After the 17 physicians completed the training, we drew a new sample from their practices (498 of 1785 consecutive patients were evaluated for depression) and measured outcomes for the depressed patients.

Results. We found an effect of the training on short-term outcome, particularly for patients with a recent-onset depression. At 3-month follow-up depressed patients whose physicians had received training had less severe psychopathology and patients with recent-onset depression also showed higher levels of daily functioning than patients of the same physicians prior to the training. The patients with a recent-onset depression that was recognized by trained physicians had shorter depressive episodes, but this was not statistically significant. At 1-year follow-up, all training effects had faded away.

Conclusions. Training primary-care physicians to recognize, diagnose and manage depression can improve short-term patient outcomes, especially for patients with a recent onset of depression. Patients suffering from a recurrent or chronic depression may need more specific interventions, both for acute treatment and long-term management.

INTRODUCTION

Depressive disorders are not always recognized and correctly diagnosed by primary-care physicians, and often patients with these disorders do not receive adequate treatment (Jenkins *et al.* 1988; Pérez-Stable *et al.* 1990; Brink *et al.* 1991; Ormel *et al.* 1991; Katon *et al.* 1992; Wells *et al.* 1994; Brody *et al.* 1995; Donoghue & Tylee, 1996). Clinical guidelines and post-graduate training programmes have been developed to improve the psychiatric knowledge and mental-health-care skills of primary-care physicians (Gask *et al.* 1988; Rutz *et al.* 1990; Kaaya *et al.* 1992; Paykel & Priest, 1992; AHCPR, 1993*a*, *b*; Marwijk *et al.* 1994). The ultimate goal of these efforts to enhance physician competence and performance is to improve patient outcomes.

Recognizing a mental health problem is a necessary but not sufficient condition to im-

[1] Address for correspondence: Dr B. G. Tiemens, Department of Psychiatry, University of Groningen, Postbus 30.001, 9700 RB Groningen, The Netherlands.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

proving outcome. Studies that examined the association between recognition and outcome have shown inconsistent results (Johnson & Goldberg, 1976; Zung *et al*. 1983; Schulberg *et al*. 1987; Ormel *et al*. 1990; Dowrick & Buchan, 1995; Simon & VonKorff, 1995; Tiemens *et al*. 1996). Studies that included only patients with new episodes and studies in which recognition consequently was followed by treatment, showed an association between recognition and better outcome (Johnston & Goldberg, 1976; Zung *et al*. 1983; Ormel *et al*. 1990), the other studies did not (Schulberg *et al*. 1987; Dowrick & Buchan, 1995; Simon & VonKorff, 1995; Tiemens *et al*. 1996). This suggests that early recognition, adequate diagnosis and treatment all are necessary to improve patient outcome.

Developing clinical guidelines and strategies to improve quality of care is a first step in changing physician performance, but these guidelines and strategies must also be implemented in everyday clinical practice. Previous studies showed that training can increase the sensitivity of primary-care physicians to mental-health problems, and that training can bring management regimens for patients with depression into concordance with clinical guidelines (Gask *et al*. 1988; Rand *et al*. 1988; Rutz *et al*. 1990; Kaaya *et al*. 1992; Roter *et al*. 1995; Lin *et al*. 1997). However, some studies showed that these positive effects on physician performance diminished over time (Rutz *et al*. 1992; Lin *et al*. 1997). The positive effects of training on patient outcomes were small or inconsistent (Roter *et al*. 1995; Lin *et al*. 1997).

Implementing new guidelines effectively requires well-designed programmes that include: (1) individual instruction; (2) practice-based reinforcing strategies; (3) peer review and feedback; and (4) reminders (Wensing & Grol, 1994; Davies *et al*. 1995). The training programmes of the studied mentioned above either focused on implementing specific diagnostic and management guidelines for depression, mainly provided by lectures (Rand *et al*. 1988; Rutz *et al*. 1990; Lin *et al*. 1997), or offered a more general clinical approach by training physicians in general communication skills (Gask *et al*. 1988; Kaaya *et al*. 1992; Roter *et al*. 1995). The programme of Lin *et al*. (1997) also included consultation sessions with a psychiatrist. The intensity of the training programmes varied from two 4-hour sessions (Roter *et al*. 1995) to eighteen 2-hour sessions (Gask *et al*. 1988).

We developed a training programme that offered both specific diagnostic and management guidelines for depression and general training in clinical approaches to mental illness. Different strategies were used, including lectures as well as practice-based hands-on-learning and feedback. In the current study we evaluated the effects of this comprehensive, 20-hour training programme for primary-care physicians (Jenner *et al*. 1995).

Two groups of physicians were trained. Van Os *et al*. (1999) described the effects of the training on physicians' performance. Recognition and correct diagnosis of depression improved after the training, but not significantly. The proportion of depressed patients who were recognized as having psychological problems, increased from 71% pre-training to 77% post-training, the proportion of patients who were diagnosed as depressed increased from 40% to 48%. The proportion of patients diagnosed as depressed by the primary-care physician that received treatment according to the trained clinical guidelines, however, increased significantly. The proportion of depressed patients getting prescriptions for antidepressants increased from 24% pre-training to 40% post-training; and the proportion of prescriptions that were for modern antidepressants (SSRIs) increased from 31% to 76%. The first trained group of physicians showed more improvement than the second group, probably due to the better pre-training performance level of the second group.

In addition to this measured effect on treatment, the training may have had an effect on management strategies that were not measured, such as providing psycho-education, and the use of motivational techniques and strategies to improve compliance.

In this paper we evaluate the effect of physician training on patient outcomes. We hypothesized that compared to depressed patients seen by untrained physicians, patients seen by trained physicians would: (1) have less severe psychopathology at follow-up; (2) recover faster (have shorter episodes of depression); and (3) return to normal daily functioning more quickly.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

METHOD

Design

We evaluated effect of training with a pre–post design. In the pre-training phase, a sample was drawn from all consecutive attenders, of the participating physicians, and examined for the presence of depression. Outcome was measured after 3 months and 1 year. Subsequently, the primary-care physicians were trained. In the post-training phase a new sample was examined and followed for 1 year.

Study setting and selection of physicians

The study was carried out in 10 primary-care practices (17 primary-care physicians) in the northern part of The Netherlands. In The Netherlands most primary-care physicians work singly in solo or double practices. The doctors assistant mainly does administrative work. For all physicians in the study, out-patient and in-patient facilities were within a distance of 25 kilometres. Access to the mental health care facilities is generally easy.

The physicians were recruited in two groups: the first group (group A) consisted of 10, and the second group (group B) of 7 primary-care physicians.

The group A physicians were selected according to the following two criteria: (1) the physician had a large practice or participated in a double or group practice; and (2) the practices and their patients had to be easy to reach by public transport. The total group had to include rural, semi-rural as well as urban practices. Of the 12 practices that met these criteria, the physicians of six practices were willing and able to participate. Reasons for physicians declining to participate were: participation in other research; illness of a physician in the practice; reorganization of the practice; no place for the research assistant to see patients; not willing to participate. The six participating practices were fairly representative of the Dutch primary-care system; two were rural (two double practices), two semi-rural (one solo, one double) and two were urban (one solo, one double).

The second group, group B, was a convenience sample of physicians who responded positively to an invitation to participate in the study. This sample consisted of physicians who were interested in, or previously had been involved in, academic psychiatric projects. Three of the seven physicians worked in a rural group practice, and four were from urban group or solo practices.

For participation in the study and visiting the training sessions the physicians were paid a small amount of money.

Intervention

The programme

The training (Jenner *et al.* 1995) consisted of eight sessions of $2\frac{1}{2}$ hours each. The main focus of the training was on depression. However, depression is often accompanied by co-morbid disorders or somatic complaints. Therefore, we also included sessions about other common mental problems. The training for group A and B physicians was slightly different, because we adapted the training to the comments of the physicians in group A. The training of the group A physicians included sessions about somatization, sleeping problems and chronic complaining behaviour, plus an introduction and booster session. In the training for group B the introduction and booster sessions were replaced by two sessions about anxiety disorders.

Teaching principles

To combine experience from primary care and specialist practice, the training programme was given by a primary-care physician (K. v.d. Meer) and a psychiatrist (J. Jenner).

Each session had a similar structure: charting of 'normal' practices and difficulties by the primary-care physician trainer; a short lecture by the psychiatrist trainer; illustration with video-taped consultations followed by discussion; introduction of guidelines and protocols for screening, diagnosis or interventions; practicing using various forms of hands-on learning (e.g. role-playing), either in small groups or plenary; evaluation. The videos consisted of examples of proper and less proper consultations. Various forms of role-playing were used: the assignment could be to role-play a doctor–patient situation on basis of a case-vignette description; to finish a video-taped consultation; or to practice working with the protocols.

Specific management guidelines

The basic strategy of the course was a so-called

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

'multi-track approach'. The strategy induces the physician to consider a psychological as well as a physical explanation for a patient's complaint, if the nature of the aetiology is not obvious. As part of this approach the somatic cues for depression were discussed in detail.

The physicians were encouraged to use a short, structured screening instrument, to determine whether further assessment of depression or anxiety is required (Tiemens *et al.* 1995). Other tools offered in the training on plasticized cards designed for daily use, were a 'symptom diagram' to facilitate diagnosis of major depression based on ICD-10 criteria for depressive episodes (World Health Organization, 1992), treatment guidelines, and medication protocols. For major depression, antidepressant medication and continuing care, including general support, counselling, psychoeducation, and planned follow-up visits, were recommended. SSRIs were recommended as first choice antidepressants because, compared to tricyclics, they are saver for use in primary care and from these antidepressants fewer side-effects and therefore better compliance could be expected. Prescription of sedatives was discouraged.

General clinical approach

The training included some general clinical management principles. The physicians were taught how to reattribute somatic complaints to psychological or psychiatric problems (Kaaya *et al.* 1992), and how to manage patients with chronic complaining behaviour. They practiced strategies for improving compliance, and techniques for motivating patients for referral to mental health care.

Furthermore, communication skills, e.g. making good eye contact, picking up verbal and non-verbal cues, timing of open and closed questions, were stressed. Such skills have been shown to facilitate detection, diagnosis, and management of psychological disorders positively (Goldberg *et al.* 1993).

Sampling and datacollection

The total data collection took place from September 1991 until March 1996. In each practice, patient enrolment was carried out during office hours over a period of 3 to 6 weeks. To prevent seasonal influences, such as winter depression, pre- and post-training samples were drawn in the same months of the year for each physician.

The pre-training sampling of the patients of the physicians of group B started 2 years later than the sampling in the group A practices, and the length of the period between the pre- and post-training measurement of the group B samples was half the length of the period between the two measurements of the group A samples.

Patients were sampled by a two-stage sampling procedure. In the first stage, consecutive patients 18–65 years completed the 12-item version of the General Health Questionnaire (Goldberg & Williams, 1988; Koeter & Ormel, 1991) while waiting to see the primary-care physician. In the second stage, patients were selected by stratified random sampling from three GHQ-strata; 10% of the patients with low scores (0,1); 35% of patients with medium scores (2–4); and 100% of the patients with high scores ($\geqslant 5$). Second-stage baseline assessment was carried out by a trained research worker within 2 weeks of screening. This was a face-to-face interview at the patient's home for the patients of the group A physicians. The patients of the group B physicians were interviewed by telephone using the same interview. After complete description of the study to the patients, written informed consent was obtained.

All patients exceeding a threshold of minimum number of current symptoms in one or more of the sections of the Composite International Diagnostic Interview-Primary Health Care Version (CIDI-PHC), were asked to participate in a 3-month and 1-year follow-up. The 3-month follow-up included self-report questionnaires. The 1-year follow-up consisted of a face-to-face or telephone interview.

Measures

The physician, seeing the patient directly after the screening, completed a Physician's Encounter Form, which records severity and diagnosis of psychological disorders. The primary-care physicians were not informed about of the results of the screening.

The second stage baseline interview included: the Composite International Diagnostic Interview-Primary Health Care Version (CIDI-PHC); 51 items of the Symptom Checklist SCL-(SCL-90); and the Brief Disability Questionnaire

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

(BDQ). The CIDI-PHC is a structured diagnostic interview developed by the World Health Organization and was adapted to the demands of the WHO primary care study Psychological Problems in General Health Care (World Health Organization, 1990; Üstün & Sartorius, 1995). The CIDI-PHC allows the assessment of the clinical state of the respondent in the last month. For the telephone interviews of the patients of the group B physicians we only used the last-month assessment. With the CIDI-PHC, diagnoses can be assessed according to ICD-10 research diagnostic criteria. For this study we used the sections on depression (ICD-10 current depression F32) and generalized anxiety disorder (ICD-10 F41–1). To measure severity, we used 51 items of the SCL-90 (SCL-51), including items of depression, anxiety, sleeping problems and somatic complaints (Derogatis *et al.* 1973; Arrindell & Ettema 1986). The average item score (range 1–5) over all 51 items or over the depression scale will be reported. The BDQ is a self-report questionnaire and was composed for the WHO primary care study (VonKorff *et al.* 1996). It includes six items concerning physical limitations from the Medical Outcome Survey Short Form 36 (MOS SF-36 (Stewart *et al.* 1988)), giving the BDQMOS score (range 0–12), five questions concerning daily functioning, motivation for work, personal efficiency and deterioration in social relations, giving the BDQ-disability (BDQDIS) score (range 0–10), and the number of activity limitation days in the prior month (BDQALD).

The self-report questionnaires at 3-month follow-up included the SCL-51 and the BDQ. The 1-year follow-up was identical with the second-stage baseline assessment. Additional information was obtained on the course of depression over the follow-up period.

Outcome variables, effect modifiers and confounding variables

Outcome variables

Episode duration was based on time to remission, as assessed by the 1-year follow-up interview. Recovery was defined as not meeting criteria for ICD-10 depression at 1-year follow-up.

Because of the skewness of the total scores of the SCL-51, the SCL depression scale (16 items), and of the BDQDIS, to examine the course of symptom severity and disability, analyses of difference scores were conducted in stead of repeated measures analyses. The variable 'activity limitation days' was also skewed; at baseline 27% of depressed patients reported no activity limitation days. In addition the variance was very high. Therefore, to examine change in activity limitation days we only included patients with any activity limitation days present at baseline. We defined the patient as recovered at 3-month or 1-year follow-up if the patient reported that he or she did not have activity limitation days in the prior month.

Effect modifiers

As treatment, psycho-education and other management strategies require recognition of the depression, an effect of the training can be expected mainly among recognized patients. We defined the depression of a patient as 'recognized by the physician', if the physician reported at the Physician's Encounter Form that the patient had a mental health problem. Furthermore, we expected less impact of training on chronically depressed patients, whose course is more difficult to change, especially with an intervention at the primary-care physician's level. The onset of depression was dichotomized. We coded onset as 'recent' if it had started within 12 months prior to baseline interview and the patient had no other episodes in the year preceding the interview. When the depressive episode had started more than 12 months before the baseline interview, or if multiple episodes had occurred in the year preceding the interview, we coded the onset as 'distant'.

Confounding variables

Other possible confounders for which we controlled in the analyses were sex, age, education level, severity of depression at baseline, comorbidity of generalized anxiety, physical limitations and the presence of a chronic disorder.

Statistical analyses

The effect of the training is examined and presented in two steps: stratified analyses and multivariate analyses. In the first step we performed univariate stratified pre–post comparisons with ANOVA for change in SCL and BDQDIS scores, Kaplan-Meier event history

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Table 1. *Pre-training and post-training response rates of the total baseline samples and follow-up response rates among patients with an ICD-10 depression, for patients of group A and B physicians*

	Group A physicians						Group B physicians					
	Pre-training		Post-training		Pre *v.* Post		Pre-training		Post-training		Pre v. Post	
	N	(%)	*N*	(%)	χ^2	*P*	*N*	(%)	*N*	(%)	χ^2	*P*
First stage												
Patients approached	1171		1195				663		590			
↓												
Patients completed GHQ	1126	(96·2)	1161	(97·2)	1·82	0·18	652	(98·3)	563	(95·4)	9·03	0·003
Second stage												
Patients sampled for interview	433		436				276		261			
↓												
Patients completed interview	300	(69·3)	295	(67·7)	0·27	0·61	218	(79·0)	203	(77·8)	0·12	0·73
↓												
Patients with an ICD-10 depression (F32/F33)	93	(31·0)	78	(26·4)	1·51	0·22	86	(39·4)	77	(37·9)	0·10	0·75
3-months follow-up completed	68	(73·1)	58	(74·4)	0·03	0·85	68	(79·1)	58	(75·3)	0·32	0·57
12-months follow-up completed	72	(77·4)	57	(73·1)	0·43	0·51	68	(79·1)	67	(87·0)	1·80	0·18
At least one follow-up completed	79	(84·9)	66	(84·6)	0·004	0·95	77	(89·5)	69	(89·6)	0·0003	0·99

analyses for episode duration, and logistic regression for proportions of patients recovered from diagnosis and free of activity limitation days. After comparing the whole pre-training sample of depressed patients *versus* the post-training sample of depressed patients, we compared the pre-training recognized patients with the post-training recognized patients, and then we stratified further to the patients with a recognized recent-onset depression pre- and post-training. In step 2 we controlled for possible confounders in multivariate analyses: ANCOVA; Cox proportional hazards regressions analysis; and multiple logistic regression.

After these two steps we examined by mixed model analyses (Goldstein, 1995) whether the results were confounded by variance among the physicians. Because we have a nested design, i.e. we sampled patients from doctors, the patients of one physician are not completely independent. This may decrease the variance among patients and consequently increases the probability of a type I error. Because, however, in these analyses, after adding the fixed effects (covariates) in each model the variance at physician level was 0, we do not present the results of these analyses.

The analyses on episode duration and recovery after 1 year included patients with an ICD-10 depressive disorder who completed the 1-year follow-up irrespective of completion of the 3-month follow-up. The analyses on the course of SCL and BDQ scores included only patients who had completed both follow-ups.

We used SPSS for Windows for all but the mixed model analyses (Norušis, 1993). For the mixed model analyses we used VARCL software (Longford, 1993). All significance tests are two-sided.

RESULTS

Study samples and patient characteristics

Table 1 shows the sampling procedure and pre- and post-training response rates for the patients of the group A physicians and group B physicians. The pre- and post-response rates differed slightly (98 % *v.* 95 %) among the group B patients in the first-stage sampling.

Among depressed patients the follow-up response rates were similar for the pre- and post-training samples. When group A and group B patients were pooled, 87·2 % of the pre-training depressed patients and 87·1 % of the post-training patients completed at least one follow-up ($\chi^2 = 0{\cdot}0002$, df = 1, $P = 0{\cdot}99$). Both follow-ups were completed by 120 (67·0 %) pre-training patients and 105 (67·7 %) post-training patients ($\chi^2 = 0{\cdot}02$, df = 1, $P = 0{\cdot}89$). Compared with the patients who completed one or two follow-ups, the patients who missed both follow-ups had less education (mean number of years of

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Table 2. *Baseline characteristics of patients with an ICD-10 depressive disorder who completed at least one follow-up, pre- and post-training (physicians pooled)*

Characteristics	Pre ($N = 156$)		Post ($N = 135$)		Pre v. Post χ^2/t	df	P
Female (%)	73·1		68·1		$\chi^2 = 0{\cdot}85$	1	0·36
Age; mean (±S.D.)	40·5	(±11·6)	38·5	(±10·4)	$t = 1{\cdot}49$	289	0·14
Employed (%)	54·8		56·3		$\chi^2 = 0{\cdot}06$	1	0·80
Years education; mean (±S.D.)	11·4	(±3·8)	12·3	(±3·5)	$t = 1{\cdot}99$	287	0·05
Chronic disorder (%)	29·0		23·0		$\chi^2 = 1{\cdot}37$	1	0·24
Self-rated overall health; mean (±S.D.)[1]	3·7	(±0·8)	3·6	(±1·1)	$t = 0{\cdot}49$	273	0·62
SCL-51[2]; mean (±S.D.)	2·41	(±0·72)	2·46	(±0·74)	$t = 0{\cdot}57$	251	0·57
SCL-depression scale (16 items) mean (±S.D.)	2·66	(±0·82)	2·72	(±0·86)	$t = 0{\cdot}57$	251	0·57
Number of CIDI depression symptoms; mean (±S.D.)	10·6	(±4·3)	10·3	(±4·3)	$t = 0{\cdot}56$	289	0·57
Number of CIDI anxiety symptoms; mean (±S.D.)	12·0	(±5·2)	11·5	(±5·0)	$t = 0{\cdot}80$	289	0·43
BDQ sum MOS-score[3]; mean (±S.D.)	4·5	(±3·4)	4·3	(±3·3)	$t = 0{\cdot}50$	278	0·62
BDQ DIS score[4]; mean (±S.D.)	4·8	(±2·6)	5·1	(±2·4)	$t = 0{\cdot}80$	278	0·43
BDQ number activity limitation days; mean (±S.D.)	11·2	(±10·7)	10·6	(±11·0)	$t = 0{\cdot}45$	273	0·65

[1] Rating on a 5-point scale from 1 = excellent to 5 = poor.
[2] SCL-51 is 51-item version of the SCL-90.
[3] BDQ = Brief Disability Questionnaire; MOS score is the total score on the six MOS-items, range 0–12.
[4] BDQ DIS score is the total score on the five items concerning daily functioning, range 0–10.

education was 10·4 (S.D. = 3·7) *v* 11·8 (S.D. = 3·7); $t = 2{\cdot}33$, df = 329, $P = 0{\cdot}02$). The respondents and non-respondents did not differ statistical significantly on other demographic or clinical (like severity of depression) variables.

Pre-training, depressed patients with at least one follow-up were similar to depressed patients post-training on all demographic and clinical variables, except years of education (Table 2).

Improvement

Table 3 shows the results of the univariate stratified analyses. At 3-month follow-up, but not at 1-year follow-up, the post-training depressed patients were more improved on the SCL-51 than pre-training. This indicates an effect of the training on short-term outcome.

In a multivariate analysis the change after 3 months in SCL-51 score, is adjusted for the independent variables (training, recognition, onset) and covariates (sex, age, education, baseline severity as measured by number of CIDI depression symptoms, co-morbid generalized anxiety and presence of chronic disorder). This ANCOVA analysis revealed an adjusted pre-training change score of 0·25 and an adjusted post-training change score of 0·41. Training was a significant predictor of 3-month change in SCL-51 score ($F = 5{\cdot}32$, df = 1, 175, $P = 0{\cdot}02$). Onset was not a statistically significant modifier, but the interaction with training was strong ($F = 2{\cdot}71$, df = 1, 175, $P = 0{\cdot}1$).

At 1-year follow-up we found no differences in SCL-51 change score between the pre-training and post-training groups. The adjusted change in SCL-51 score was 0·55 pre-training and 0·59 post-training ($F = 0{\cdot}65$, df = 1, 175, $P = 0{\cdot}42$). However, recognition ($F = 4{\cdot}87$, df = 1, 175, $P = 0{\cdot}03$), recency of onset ($F = 4{\cdot}73$, df = 1, 175, $P = 0{\cdot}03$) and the interaction of recognition and recency of onset ($F = 4{\cdot}18$, df = 1, 175, $P = 0{\cdot}04$) predicted a larger change in SCL-51 score, indicating a better outcome.

Analysing the SCL depression scale separately, we found an adjusted change in depression scale score after 3 months of 0·35 pre-training and of 0·57 post-training. The ANCOVA showed a main effect of training ($F = 5{\cdot}89$, df = 1, 175, $P = 0{\cdot}02$). At 1-year follow-up there was no pre–post difference.

Recovery

The event history analyses (Kaplan-Meier) on time to recovery showed a slightly quicker recovery in the post-training group of patients, but this difference was not statistically significant (Fig. 1*a*). Stratifying to the patients whose depression was recognized by the physicians, the pre–post difference in recovery time was somewhat larger (Fig. 1*b*), and the Kaplan-Meier curves of the patients with a recognized recent onset depression, despite the small numbers, showed a difference with $P = 0{\cdot}08$ (Fig. 1*c*). Those patients whose physician had been trained

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Table 3. *Change in SCL-51 score after 3 months and 1 year, pre-training and post-training, among all depressed patients, the recognized patients and the recognized patients with a recent onset depression*

	All depressed patients				Recognized patients				Recognized patients with recent onset			
SCL-51	Pre-training ($N = 89$) Mean	(S.D.)	Post-training ($N = 87$) Mean	(S.D.)	Statistic t (df = 174)	Pre-training ($N = 63$) Mean	(S.D.)	Post-training ($N = 70$) Mean	(S.D.)	Statistic t (df = 131)	Pre-training ($N = 34$) Mean	(S.D.)	Post-training ($N = 37$) Mean	(S.D.)	Statistic t (df = 69)
Baseline, mean (S.D.)	2·35	(0·73)	2·47	(0·78)	1·08	2·52	(0·77)	2·55	(0·75)	0·24	2·39	(0·77)	2·35	(0·63)	0·29
Difference at 3-month follow-up mean (S.D.)	0·23	(0·62)	0·44	(0·62)	2·24**	0·31	(0·63)	0·46	(0·64)	1·39	0·33	(0·73)	0·61	(0·67)	1·67*
Difference at 1-year follow-up mean (S.D.)	0·53	(0·73)	0·62	(0·64)	0·78	0·60	(0·81)	0·67	(0·66)	0·57	0·71	(0·82)	0·86	(0·64)	0·86

* $P < 0{\cdot}1$; ** $P < 0{\cdot}05$.

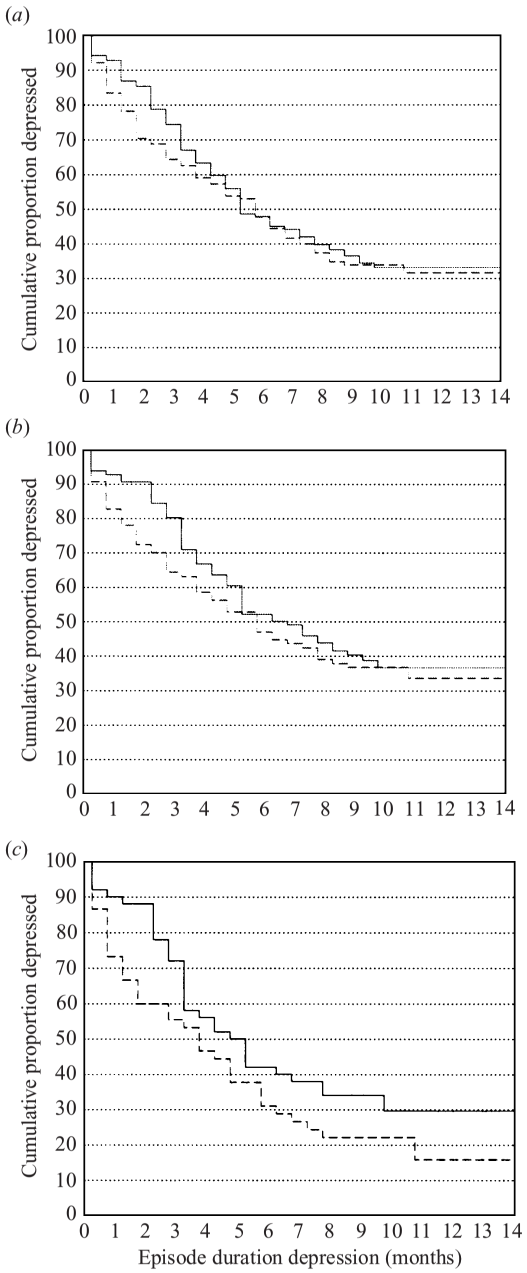


FIG. 1. Kaplan-Meier curves for time to recovery from depression: stratified analyses. (*a*). Patients with depression (——, pre-training $N = 136$; – – –, post-training $N = 115$), Breslow test $P = 0{\cdot}33$. (*b*) Patients with recognized depression (——, pre-training $N = 96$; – – –, post-training $N = 87$), Breslow test $P = 0{\cdot}19$. (*c*) Patients with recognized recent onset depression (——, pre-training $N = 50$; – – –, post-training $N = 45$), Breslow test $P = 0{\cdot}08$.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Table 4. *Change in BDQDIS score after 3 months and 1 year, pre-training and post-training, among all depressed patients, the recognized patients and the recognized patients with a recent onset depression*

	All depressed patients				Recognized patients				Recognized patients with recent onset					
BDQDIS	Pre-training ($N = 107$) Mean	(S.D.)	Post-training ($N = 92$) Mean	(S.D.)	Statistic t (df = 197)		Pre-training ($N = 76$) Mean	(S.D.)	Post-training ($N = 73$) Mean	(S.D.)	Statistic t (df = 131)		Pre-training ($N = 35$) Mean	(S.D.)	Post-training ($N = 36$) Mean	(S.D.)	Statistic t (df = 69)	
Baseline, mean (S.D.)	4·73	(2·56)	5·09	(2·43)	1·01		5·00	(2·60)	5·21	(2·47)	0·49		5·03	(2·41)	5·28	(2·55)	0·42	
Difference at 3-month follow-up mean (S.D.)	0·97	(2·66)	1·42	(2·28)	1·28		1·13	(2·56)	1·48	(2·36)	0·86		0·91	(2·77)	2·06	(1·97)	2·01	**
Difference at 1-year follow-up mean (S.D.)	1·97	(3·63)	2·14	(3·04)	0·35		2·17	(3·47)	2·48	(2·91)	0·59		2·17	(3·46)	3·31	(2·92)	1·50	

** $P < 0{\cdot}05$.

seemed to recover more rapidly than patients of the same physicians prior to training.

In a Cox regression including all depressed patients who completed the 1-year follow-up, we examined the effects on episode duration, of training (pre *versus* post), recognition, and recency of onset. We controlled for sex, age, education level, severity of depression (number of CIDI depression symptoms), co-morbidity of generalized anxiety and the presence of a chronic disorder. As expected from the Kaplan-Meier curves we did not find a main effect of training. The hazard ratio (HR) for interaction of training with onset was 1·82 (95% CI 0·94–3·54). Although not statistically significant, the HR suggests that the training was more effective for the patients with recent onset. Significant confounders were higher level of education (HR = 1·08 per education year; 95% CI 1·03–1·12) and lower number of depression symptoms (HR = 0·91, per symptom; 95% CI 0·87–0·95).

The Kaplan-Meier curves show that at the end of the follow-up period, the number of patients that had been recovered was not higher after the training. This finding is confirmed by the 1-year follow-up CIDI-PHC results. The proportion of the depressed patients who did not meet criteria for an ICD-10 depressive disorder after 1 year, was 69·3% pre-training and 71·8% post-training (OR = 1·13; 95% CI 0·64–1·99). Among the patients whose depression was recognized by the physician, this was 67·3% pre-training and 73·7% post-training (OR = 1·36; 95% CI, 0·70–2·66) and among the recognized patients with a recent-onset depression it was 80·0% pre-training *versus* 87·5% post-training (OR = 1·75; 95% CI 0·52–6·40). In a multivariate logistic regression with recovery as the dependent variable, testing the same model as with the Cox regression, we found no effect of training. The odds ratio for being recovered after a year for the patients with recent onset *versus* distant onset was 3·4 (95% CI 1·41–8·32).

Although these results on episode duration do not show a statistically significant effect of the training, the direction of the results corroborate the findings with the SCL of an effect of the training on short-term outcome or in other words a quicker improvement after the training.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Disability

For changes in daily functioning we examined: (1) the change in BDQDIS score; and (2) the proportion of patients without activity limitation days at 3-month and 1-year follow-ups. The pre–post difference in change in BDQDIS score at 3-month and 1-year follow-up was largest among recognized patients with a recent-onset depression (univariate stratified analyses in Table 4).

We analysed the change in BDQDIS score after 3 months, adjusting for independent variables (training, recognition, onset) and covariates (sex, age, education, presence of chronic disorder, baseline severity, co-morbid anxiety and physical limitations). This ANCOVA analysis revealed an adjusted pre-training change score of 1·03 and an adjusted post-training change score of 1·35. Here we found an interaction effect of training with onset ($F = 5{\cdot}30$, df $= 1, 198$, $P = 0{\cdot}02$). This means that patients post-training with recent onset depression improved more on daily functioning. After a year we found no pre–post difference.

At baseline interview 140 patients with an ICD-10 depression (78 pre-training and 62 post-training) reported activity limitation days in the prior month. After 3 months, 22 of the initial 78 patients (28·2%) reported pre-training no activity limitation days and 20 of the initial 62 patients (32·3%) post-training had no activity limitation days (OR = 1·21; 95% CI 0·55–2·66). Among the recognized patients ($N = 104$), these proportions were 19·6% pre-training and 33·3% post-training (OR = 2·05; 95% CI 0·77–5·55), and among the recognized patients with a recent onset depression ($N = 53$) these proportions were 13·8% pre-training and 41·7% post-training (OR = 4·46; 95% CI 1·01–22·60). In a logistic regression we found an odds ratio for interaction of the training with recognition of 5·96 (95% CI 0·99–35·70), indicating that recognized patients in the post-training group were more likely to be free from activity limitation days after 3 months than those in the pre-training group. Patients with higher number of depression symptoms were less likely to be free from activity limitation days after 3 months (OR = 0·88 per symptom; 95% CI 0·78–0·99).

At 1-year follow-up no pre–post difference was found.

Differences between the two groups of physicians

We did all analyses for the two groups of physicians separately. In some analyses the effect of the training on patient outcomes was stronger in group A compared to group B physicians, but in no case were the effects in opposite directions.

DISCUSSION

The results of this study show that a comprehensive, multi-faceted training for primary-care physicians can improve patient outcomes, particularly for patients with a recent-onset depression. Although we did not find a statistically significant pre–post difference in all outcome variables, they all point in the direction of an effect of the training on short-term outcome: the depressed patients of the trained physicians were more improved at 3-month follow-up, but not at 1-year follow-up; the patients with recent onset depression recovered faster, i.e. had shorter episodes, and were more improved in daily functioning at 3-month follow-up, but not at 1-year follow-up.

The effect of the training might have been strengthened or weakened by three other kinds of effect; (1) a time effect strengthening the training effect, which should have benefited in particular the group A physicians because of a longer pre–post interval; (2) an extinction effect weakening the training effect which also should have been stronger for the group A physicians; and (3) the effect of differences in pre-training performance between the physicians.

Time-effect

The main limitation of this study is the pre–post design. Although in a randomized clinical trial it is difficult to reach a balanced randomization with less than 60 physicians (and we had only 17), a pre–post design has the risk of a time effect. A time effect in the study period 1991–1996 is likely, especially in the group A physicians with up to 3 years between the pre- and post-training baseline assessments. An ongoing stream of information about the effectiveness of the SSRIs not only reached physicians, but their patients as well. Prescriptions for antidepressants increased rapidly over this period of time

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

(*Geneesmiddelenbulletin*, 1995). As we found more post-training patients had prescriptions for antidepressant (Van Os *et al*. 1999), we cannot exclude that the training effect reflects this general development.

Extinction effect

Although time was in favour of the group A physicians, the risk of extinction of the training effect was also higher in this group. Lin *et al*. (1997) found that the effect of the training extinguished immediately after the training had stopped. Rutz *et al*. (1992) recommended booster sessions to maintain the training effect. In our study the period between training and post-training assessment was longer for the group A physicians than for group B physicians, but we found the same or stronger results among group A physician's patients. The practical tools, as plasticized cards with the diagnostic and treatment guidelines, may have facilitated the process of practicing the new skills after the training was terminated.

Pre-training performance level

Another limitation of the study is the way in which the primary-care physicians were selected. The physicians in the group B were from a selective group with a better baseline performance level. Therefore, it may have been more difficult to achieve improvement in group B. The group A physicians, who were more representative of Dutch primary-care physicians, improved more at process-of-care variables after the training than group B physicians (Van Os *et al*. 1999). We also found in some, but not all, analyses a stronger effect of the training on outcome among the patients of the group A physicians.

Several real training aspects may be responsible for the effect of the training on outcome. Unfortunately, in this study we did not have the resources to measure what actually happened during the patient's visit to the doctor pre- and post-training. In the training we recommended that physicians combine prescription of antidepressants with continuing care, support, counselling and psycho-education. Motivational techniques and strategies to improve compliance were also practiced in the training. Lin *et al*. (1995) found that psycho-education increased patients' adherence to medication. We found that the combination of prescribing antidepressants and support/counselling increased after the training (Van Os *et al*. 1999). Better psycho-education may have been a component of this general support, which may have enhanced the effectiveness of antidepressant medication by better adherence.

What are the clinical implications of the findings of this study? From a clinical point of view, the effects of the training are not very large. First, the effect of training on the course of depression seems be limited to speeding up recovery, and does not seem to affect long-term outcome. However, because most depressed patients are improved after a year the power to detect a pre–post training difference at 1-year follow-up is very small. Secondly, although many patients improve, the long-term outcome is not very favourable, they do not fully recover, even post-training. The mean SCL symptom score at 1-year follow-up in the post-training group was similar to the highest norm score in the Dutch population (Arrindell & Ettema 1986).

The effect of the training, however, was different for subgroups of patients. The outcome of patients whose depression started in the year prior to the baseline interview was better compared to patients whose depression had a more chronic course. This is consistent with the finding of Scott *et al*. (1992) that early intervention improves prognosis. Therefore, if trained physicians were to recognize patients' depression at an earlier stage this may reduce the number of chronic patients.

For the patients whose depression has a more chronic course the training for primary-care physicians does not seem sufficient. These patients require more care than the treatment modalities in which we have trained the physicians. An option may be a combination of acute treatment with long-term management as described in the population-based care model of Katon *et al*. (1997). This model consists of four steps to decrease the prevalence of depression: (1) improving recognition and accuracy of diagnosis; (2) improving acute phase management of depression; (3) preventing chronicity among cases beginning the transition from acute to persistent of chronic illness by augmenting mental health services; and (4) preventing relapse/recurrence among patients at high risk

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

of relapse in order to reduce the mean number of episodes over the life-span. The training of primary-care physicians may be crucial for the implementation of this model, not only for the first two steps, but also by improving practice, alertness, and decision making in the other steps. However, as a recent study again shows (Goldberg *et al*. 1998) that low-intensity programmes have little effect on physician practice or patient outcomes, such training should be multi-faceted and intensive.

We gratefully acknowledge the participating primary-care physicians, the research workers who interviewed the patients, L. Carter for her comments on an earlier version of this paper, A. van den Bos for her help with the mixed model analyses, C. Rutter for statistical advice, and H. Hallie for the data entry.

REFERENCES

Agency for Health Care Policy and Research (1993*a*). *Depression in Primary Care: Volume 1. Detection and Diagnosis*. Clinical Practice Guideline Number 5: Rockville, Maryland.

Agency for Health Care Policy and Research (1993*b*). *Depression in Primary Care: Volume 2. Treatment of Major Depression*. Clinical Practice Guideline Number 5: Rockville, Maryland.

Arrindell, W. A. & Ettema, H. (1986). *SCL-90: Handleiding bij een Multidimensionele Psychopathologie Indicator*. Swets & Zeitlinger: Lisse.

Brink, W. van den, Leenstra, A., Ormel, J. & Willige, G. van den. (1991). Mental health intervention programs in primary care: their scientific basis. *Journal of Affective Disorders* **21**, 273–284.

Brody, D. S., Thompson, T. L., Larson, D. B., Ford, E. D., Katon, W. J. & Magruder, M. (1995). Recognizing and managing depression in primary care. *General Hospital Psychiatry* **17**, 93–107.

Davis, D. A., Thomson, M. A., Oxman, A. D. & Haynes, R. B. (1995). Changing physician performance. A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association* **274**, 700–705.

Derogatis, L. R., Lipman, R. S. & Covi, L. (1973). SCL-90 an outpatient rating-scale: preliminary report. *Psychopharmacology Bulletin* **9**, 13–27.

Donoghue, J. M. & Tylee, A. (1996). The treatment of depression: prescribing patterns of antidepressants in primary care in the UK. *British Journal of Psychiatry* **168**, 164–168.

Dowrick, C. & Buchan, I. (1995). Twelve month outcome of depression in general practice: does detection or disclosure make a difference? *British Medical Journal* **311**, 1274–1276.

Gask, L., Goldberg, D., Lesser, A. L. & Millar, T. (1988). Improving the psychiatric skills of the general practice trainee: an evaluation of a group training course. *Medical Education* **22**, 132–138.

Geneesmiddelenbulletin (1995). Koste-wat-het-kost? Ontwikkelingen in het gebruik van antidepressiva **29**, 120.

Goldberg, D. & Williams, P. (1988). *A Users' Guide to the General Health Questionnaire*. NFER-Nelson: Windsor.

Goldberg, D. P., Jenkins, L., Millar, T. & Faragher, E. B. (1993). The ability of trainee general practitioners to identify psychological distress among their patients. *Psychological Medicine* **23**, 185–193.

Goldberg, H. I., Wanger, E. H., Fihn, S. D., Martin, D. P., Horowitz, C. R., Christensen, D. B., Cheadle, A. D., Diehr, P. & Simon, G. (1998). A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *Journal of Quality Improvement* **24**, 130–142.

Goldstein, H. (1995). *Multilevel Statistical Models*, *2nd edn*. Edward Arnold: London.

Jenkins, R., Smeeton, N. & Shepherd, M. (1988). *Classification of Mental Disorder in Primary Care. Psychological Medicine Monograph supplement 12*. Cambridge University Press: Cambridge.

Jenner, J. A., Meer, K. van der, Os, T. W. D. P. van Brink, R. H. S. van den, Tiemens, B. G. & Ormel, J. (1995). *Psychosociale klachten in de huisartspraktijk, een cursus voor huisartsen*. INSTEL-project, Rijksuniversiteit Groningen, Van Gorcum: Assen.

Johnstone, A. & Goldberg, D. (1976). Psychiatric screening in general practice. A controlled trial. *Lancet* i, 605–609.

Kaaya, S., Goldberg, D. & Gask, L. (1992). Management of somatic presentations of psychiatric illness in general medical settings: evaluation of a new training course for general practitioners. *Medical Education* **26**, 138–144.

Katon, W., Von Korff, M., Lin, E., Bush, T. & Ormel, J. (1992). Adequacy and duration of antidepressant treatment in primary care. *Medical Care* **30**, 67–76.

Katon, W., VonKorff, M., Lin, E., Unützer, J., Simon, G., Walker, E., Ludman, E. & Bush, T. (1997). Population-based care of depression: effective disease management strategies to decrease prevalence. *General Hospital Psychiatry* **19**, 168–178.

Koeter, M. W. J. & Ormel, J. (1991). *General Health Questionnaire. Nederlandse bewerking en Handleiding*. Swets & Zeitlinger: Lisse.

Lin, E. H. B., Von Korff, M., Katon, W., Bush, T., Simon, G. E., Walker, E. & Robinson, P. (1995). The role of the primary care physician in patients' adherence to antidepressant therapy. *Medical Care* **33**, 67–74.

Lin, E., Katon, W., Simon, G., Von Korff, M., Bush, T., Rutter, C., Saunders, K. & Walker, E. (1997). Achieving guidelines for the treatment of depression in primary care: is physician education enough? *Medical Care* **35**, 831–842.

Longford, N. T. (1993). *VARCL: Software for Variance Component Analysis of Data with Nested Random Effects* (*Maximum Likelihood*). *Manual*. ProGAMMA: Groningen.

Marwijk, H. W. J. van, Grundmeijer, H. G. L. M., Brueren, M. M., Sigling, H. O., Stolk, J., Gelderen, M. G. van, Vintges, M., Eizenga, W. H. & Burgers, J. S. (1994). NHG-Standaard Depressie. Nederlands Huisartsen Genootschap Standaard M43. *Huisarts en Wetenschap* **7**, 482–490.

Norušis, M. J. (1993). *SPSS for Windows*, *Release 6.0*. SPSS Inc.: Chicago.

Ormel, J., Brink, W. van der, Koeter, M. W. J., Giel, R., Meer, K. van der, Willige, G. van de & Wilmink, F. W. (1990). Recognition, management and outcome of psychological disorders in primary care: a naturalistic follow-up study. *Psychological Medicine* **20**, 909–923.

Ormel, J., Koeter, M. W. J., Brink, W. van den & Willige, G. van de. (1991). Recognition management and course of anxiety and depression in general practice. *Archives of General Psychiatry* **48**, 700–706.

Paykel, E. S. & Priest, R. G. (1992). Recognition and management of depression in general practice: consensus statement. *British Medical Journal* **305**, 1198–1202.

Pérez-Stable, E. J., Miranda, J., Muñoz, R. F. & Ying, Y.-W. (1990). Depression in medical outpatients. Underrecognition and misdiagnosis. *Archives of Internal Medicine* **150**, 1083–1088.

Rand, E. H., Badger, L. W. & Coggins, D. R. (1988). Toward a resolution of contradictions. Utility of feedback from the GHQ. *General Hospital Psychiatry* **10**, 189–196.

Roter, D. L., Hall, J. A., Kern, D. E., Barker, R., Cole, K. A. & Roca, R. P. (1995). Improving physicians' interviewing skills and reducing patients' emotional distress. *Archives of Internal Medicine* **155**, 1877–1884.

Rutz, W., Knorring, von L., Wålinder, . & Wistedt, B. (1990). Effect of an educational program for general practitioners on Gotland on the pattern of prescription of psychotropic drugs. *Acta Psychiatrica Scandinavica* **82**, 399–403.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press

Rutz, W., Knorring, von L. & Wålinder, J. (1992). Long-term effects of an educational program for general practitioners given by the Swedish Committee for the Prevention and Treatment of Depression. *Acta Psychiatrica Scandinavica* **85**, 83–88.

Schulberg, H. C., McClelland, M. & Gooding, W. (1987). Six-month outcomes for medical patients with major depressive disorders. *Journal of General Internal Medicine* **2**, 312–317.

Scott, J., Eccleston, D. & Boys, R. (1992). Can we predict the persistence of depression? *British Journal of Psychiatry* **161**, 633–637.

Simon, G. E. & Von Korff, M. (1995). Recognition, management, and outcomes of depression in primary care. *Archives of Family Medicine* **4**, 99–105.

Stewart, A. L., Hays, R. D. & Ware, J. E. J. (1988). The MOS short-form general health survey: reliability and validity in a patient population. *Medical Care* **26**, 724–735.

Tiemens, B. G., Ormel, J., Brink, R. H. S. van den, Jenner, J. A., Meer, K. van der, Os, T. W. D. P. van & Smit, A. (1995). Signalering van depressie en gegeneraliseerde angst in de huisartspraktijk; de onwikkeling van een screeninginstrument. *Tidschrift voor Sociale Gezondheidszorg* **8**, 520–527.

Tiemens, B. G., Ormel, J. & Simon, G. E. (1996). Occurrence, recognition, and outcome of psychological disorders in primary care. *American Journal of Psychiatry* **153**, 636–644.

Üstün, T. B. & Sartorius, N., editors. (1995). *Mental Illness in General Health Care. An International Study*. John Wiley & Sons: New York.

Van Os, T. W. D. P., Ormel, J., Brink, R. H. S. van den, Jenner, J. A., Meer, K. van der, Tiemens, B. G., Doorn, W. van der, Smit, A. & Brin, W. van den. (1999). Improving the management of depression in primary care: the effect of a training program for general practitioners in The Netherlands. *General Hospital Psychiatry* (in the press).

Von Korff, M., Üstün, T. B., Ormel, J., Kaplan, I. & Simon, G. E. (1996). Self-report disability in an international primary care study of psychological illness. *Journal of Clinical Epidemiology* **49**, 297–304.

Wells, K. B., Katon, W., Rogers, B. & Camp, P. (1994). Use of minor tranquilizers and antidepressant medications by depressed outpatients: results from the Medical Outcomes Study. *American Journal of Psychiatry* **151**, 694–700.

Wensing, M. & Grol, R. (1994). Single and combined strategies for implementing changes in primary care: a literature review. *International Journal for Quality in Health Care* **6**, 115–132.

World Health Organization (1990). *Composite International Diagnostic Interview* (*CIDI*), *Core Version 1.0 – November 1990*. WHO: Geneva.

World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines*. WHO: Geneva.

Zung, W. W. K., Magill, M., Moore, J. T. & George, D. T. (1983). Recognition and treatment of depression in a family medicine practice. *Journal of Clinical Psychiatry* **44**, 3–6.

https://doi.org/10.1017/S0033291799008545 Published online by Cambridge University Press