Introduction

EVERY day general practitioners (GPs) see patients who present with physical symptoms for which there is no specific diagnosis and for which medicine does not provide a cure. The profusion of labels, such as ‘medically unexplained physical symptoms’, ‘somatisation’ and ‘hypochoONDria’ testifies to the prevailing diversity of conception of these symptoms. We have adopted Lipowski’s definition of somatisation as ‘a tendency to experience and communicate somatic distress and symptoms unaccounted for by pathological findings, to attribute them to physical illness and to seek medical help for them’.¹ This definition encompasses the broad spectrum of disorders that are encountered in general practice.

Somatoform disorder may be diagnosed in 20–30% of primary care patients on the basis of psychiatric standard interviews.²⁻⁵ GPs only recognise about 50% of patients who are diagnosed by psychiatrists.²,⁵ Unrecognised cases may not receive proper treatment, and are at risk of iatrogenic harm and disability during the course of ruling out physical disease.⁶,⁷

A review of controlled interventional studies on mental disorders found that 18 of 23 trials showed improvement in diagnosis, but none of the studies specifically addressed somatisation.⁸ Educational interventions have increased GPs’ rates of detection of emotional distress,⁹⁻¹¹ but failed to improve their detection of depression in a recent randomised controlled trial.¹² Studies on somatisation have shown that training may improve interviewing skills,¹³ and that implementation of new skills in daily clinical practice is feasible.¹⁴ However, interventions that target diagnostic skills alone may be insufficient. Improved recognition rates also depend on increased possibilities for treatment or management of patients.¹⁵

The aim of this intervention study was to evaluate the effect of a novel multifaceted training programme on GPs’ diagnosis of somatisation during routine clinical practice.¹⁶ We hypothesised that trained GPs would recognise a larger number of somatising patients, and that their diagnosis would show better agreement with psychiatric rating scales than those of control GPs.
Danish healthcare system is tax-financed and 98% of Danish people are listed with one general practice.

General practitioners and randomisation

GPs who were registered with the Vejle County Health Insurance were invited to participate in the study in November 1999. Inclusion criteria consisted of: participation of at least 50% of GPs from a practice, and minimum working hours 2.5 days per week. Enrolled practices were stratified according to the number of GPs per practice (1–4) and the proportion of participating GPs in relation to the total number of GPs in the practice (0.5–1.0). After the inclusion of practices was completed, practices in each stratum were allocated to intervention or control (Figure 1 and Table 1). An individual who was not involved in the study performed the randomisation by drawing non-transparent lots containing code numbers. Practices could not be blinded, but were asked not to inform patients about their grouping. All GPs received reimbursement for participating in the study.

Patients

Practice secretaries enrolled patients consecutively over a period of 13 working days in May 2000. Inclusion criteria were that patients should be aged 18–65 years and consulting about a new health problem. Patients were excluded if they had acute severe disease (n = 19), mental handicap (n = 38), if they were of non-Scandinavian descent (n = 311), if they were not listed with a participating GP (n = 53), or if participation was not possible for other reasons (for example, error in registration number or procedure, unable to read or write because of forgotten glasses or problems with arm [n = 281]) (Figure 1). Practice secretaries and GPs registered patients independently.

Sample size

The level of GP diagnoses (20%) was assessed from previous studies. Standard power calculations showed that 1733 patients had to be included in each arm to obtain a minimum required difference in diagnoses of 20% (type 1 error, 0.05; type 2 error, 0.20). On average, GPs would include 80 patients. Thus, a total sample size of 44 GPs was desired. Subsequent power analysis based on the observed cluster adjusted standard error of diagnostic difference indicated that it was necessary to double the GP sample size.

Intervention

Intervention consisted of a multifaceted educational programme on assessment, treatment, and management of...
somatisation (the TERM [The Extended Reattribution and Management] model), which is based on current theoretical knowledge and scientific data on somatisation, and adopts a cognitive-oriented approach. It is described in detail elsewhere and is summarised in Box 1. The training programme included positive criteria for somatisation and skills training in patient examination and consultation. They were offered the opportunity to take part in the training programme following completion of the trial.

### Outcome measures

Baseline characteristics of GPs were obtained from Vejle County Health Insurance and from questionnaires on GPs’ postgraduate training. All patients who were included in the study completed a screening questionnaire in the waiting room before their consultation. GPs filled in a questionnaire independently after the consultation. The results of the questionnaires were only revealed to the project leader.

Two ratings were made, namely a GP diagnosis and a psychometric assessment:

- GPs were asked to classify the main problem presented by the patient in one of five categories (Table 2). During analyses, diagnoses were dichotomised as either ‘physical disease’ or ‘somatisation’.
- For the psychometric assessment a somatisation subscale from the Hopkins symptom checklist (SCL-SOM) and a scale for illness worrying and conviction (Whiteley-7) were applied. Patients were asked about symptoms they had experienced during the past 4 weeks and responses were given on a 5-point Likert scale. Both scales were dichotomised between the second and third response categories. The cut-off points chosen for this study were 3/4 for SCL-SOM and 1/2 for Whiteley-7. Compared with a standardised interview (schedules for clinical assessment in neuropsychiatry), the sensitivity, specificity, and positive predictive values for any somatoform disorder were 0.378, 0.833, 0.586, respectively, for SCL-SOM and 0.311, 0.856, 0.574, respectively, for Whiteley-7 (Kaj Sparie Christensen, personal communication, 2002). Screening was positive if scoring was high on at least one scale.

Primary end-points were the GPs’ diagnosis of somatisation and their diagnostic agreement with rating scales.

### Statistical analysis and software

Questionnaire data were processed using TELEform formulae. Intention-to-treat analyses for GPs could not be performed, as lost GPs did not provide the necessary information. With regard to patients, information was missing for 15% of those who refused to participate in the study. Consequently, analyses were performed using complete data only. Analyses at patient level were adjusted for patient sex and clusters.

| Table 1. Baseline characteristics for participating general practitioners in Vejle County. |

<table>
<thead>
<tr>
<th></th>
<th>Participating GPs (n = 37)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention n = 20</td>
<td>Control n = 17</td>
</tr>
<tr>
<td>Male n (%)</td>
<td>15 (75.0)</td>
<td>11 (64.7)</td>
</tr>
<tr>
<td>Mean age (years) (SD)</td>
<td>48.1 (7.7)</td>
<td>47.9 (6.3)</td>
</tr>
<tr>
<td>Mean seniority as GP (SD)</td>
<td>12.0 (9.8)</td>
<td>7.6 (7.0)</td>
</tr>
<tr>
<td>Partnership practice n (%)</td>
<td>14 (70.0)</td>
<td>12 (70.6)</td>
</tr>
<tr>
<td>Urban location of practice n (%)</td>
<td>19 (95.0)</td>
<td>17 (100.0)</td>
</tr>
<tr>
<td>Median number of GPs in practice, (25–75% percentiles)</td>
<td>3 (1–4)</td>
<td>2 (1–3)</td>
</tr>
<tr>
<td>Mean number of listed patients per GP (SD)</td>
<td>1526 (291)</td>
<td>1645 (185)</td>
</tr>
<tr>
<td>Previous longer courses, n (%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6 (30.0)</td>
<td>8 (50.0)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Previous supervision, n (%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>9 (47.4)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8 (50.0)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>x<sup>2</sup> test; <sup>b</sup>Mann–Whitney U-test; <sup>c</sup>Supervision and courses of at least 3 days duration concerning communication skills and psychiatric training; <sup>d</sup>Missing information on one GP.
Although randomisation was performed at practice level, clustering within GPs was thought to be of greater importance for diagnoses than clustering within practices. The \( \chi^2 \) test was applied to dichotomous data and the \( t \)-test or Mann–Whitney U-test was applied to continuous data. We used general linear models with an identity link for ... a random-effects model. Statistical analyses were performed using STATA SE version 8.0 and SPSS version 10.0 for Windows.

### Ethics and approval

This study was approved by the Ethics Committee for Funen and Vejle County, the Data Surveillance Authority and the Scientific Research Evaluation Committee of the Danish College of General Practitioners.

### Results

A total of 27 practices including 43 GPs were enrolled and randomised to intervention (14 practices including 23 GPs) or control (13 practices including 20 GPs) (Figure 1). Three practices dropped out before intervention, and a further three were excluded because of low rates of patient enrolment. Participants who completed the study had practised family medicine for fewer years than non-participants (10.0 versus 12.8; \( P = 0.038 \)) and were more likely to be from urban areas (97.3% versus 56.0%; \( P < 0.001 \)) but otherwise did not differ significantly from non-participants on the parameters listed in Table 1. The randomised groups did not differ significantly on the parameters listed in Table 1.

In total, 15% of patients who were eligible for inclusion refused to participate in the study, and 9% were not asked in error (Figure 1). Refusers were older than participants (mean

### Table 2. General practitioners' classification of the main problem presented by the patient in the consultation with numbers of patients presented according to classification and randomisation group.

<table>
<thead>
<tr>
<th>Classification category</th>
<th>Intervention</th>
<th>Control</th>
<th>Adjusted diagnostic difference(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical disease</td>
<td>960</td>
<td>918</td>
<td>-6.2 ( \Delta% )</td>
</tr>
<tr>
<td>Probable physical disease</td>
<td>247</td>
<td>188</td>
<td>2.8 ( \Delta% )</td>
</tr>
<tr>
<td>Medically unexplained symptoms</td>
<td>109</td>
<td>48</td>
<td>4.0 ( \Delta% )</td>
</tr>
<tr>
<td>Mental illness</td>
<td>29</td>
<td>35</td>
<td>-0.8 ( \Delta% )</td>
</tr>
<tr>
<td>No physical symptoms</td>
<td>62</td>
<td>51</td>
<td>0.2 ( \Delta% )</td>
</tr>
<tr>
<td>Subtotal physical</td>
<td>(1207)</td>
<td>(1106)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)All analyses were adjusted for patient sex and clustering of patients within GPs; \(^b\)The diagnostic difference was calculated by subtracting the percentage in the control group from the percentage in the intervention group; \(^c\)Comparisons of each classification category against the sum of others.

Comparison of all classification categories in one analysis, \( P = 0.04 \). The GPs' diagnoses were dichotomised around the broken line for the diagnosis of somatisation.

### Table 3. The general practitioners' classification of the patient's main problem as 'somatisation' or 'physical disease' compared with results from a patient screening questionnaire.

<table>
<thead>
<tr>
<th>Screening questionnaire</th>
<th>Positive(^a)</th>
<th>Negative(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP + Intervention</td>
<td>98 (21.0)</td>
<td>97 (10.5)</td>
</tr>
<tr>
<td>Control</td>
<td>66 (17.6)</td>
<td>68 (8.0)</td>
</tr>
<tr>
<td>( \Delta% ) agreement(^c)</td>
<td>3.4 (-3.4 to 10.1)(^d)</td>
<td>0.327</td>
</tr>
<tr>
<td>( P)-value(^c)</td>
<td>0.161</td>
<td></td>
</tr>
</tbody>
</table>

| GP - Intervention       | 368 (79.0)    | 829 (89.5)    | 1197 |
| Control                 | 308 (82.4)    | 787 (92.0)    | 1095 |
| \( \Delta\% \) agreement\(^d\) | -2.5 (-6.1 to 1.0)\(^e\) | 0.161 |
| \( P\)-value\(^c\)       | 0.161 |

Although randomisation was performed at practice level, clustering within GPs was thought to be of greater importance for diagnoses than clustering within practices. The \( \chi^2 \) test was applied to dichotomous data and the \( t \)-test or Mann–Whitney U-test was applied to continuous data. We used general linear models with an identity link for Bernoulli family; that is, modelling the risk differences, when adjusting analyses. This was supplemented by the Wald test in a combined analysis of diagnostic categories. The intra-cluster variation was assessed by the one-way analysis of variance (ANOVA) for a random-effects model. Statistical analyses were performed using STATA SE version 8.0 and SPSS version 10.0 for Windows.

### Ethics and approval

This study was approved by the Ethics Committee for Funen and Vejle County, the Data Surveillance Authority and the Scientific Research Evaluation Committee of the Danish College of General Practitioners.

### Results

A total of 27 practices including 43 GPs were enrolled and randomised to intervention (14 practices including 23 GPs) or control (13 practices including 20 GPs) (Figure 1). Three practices dropped out before intervention, and a further three were excluded because of low rates of patient enrolment. Participants who completed the study had practised family medicine for fewer years than non-participants (10.0 versus 12.8; \( P = 0.038 \)) and were more likely to be from urban areas (97.3% versus 56.0%; \( P < 0.001 \)) but otherwise did not differ significantly from non-participants on the parameters listed in Table 1. The randomised groups did not differ significantly on the parameters listed in Table 1.

In total, 15% of patients who were eligible for inclusion refused to participate in the study, and 9% were not asked in error (Figure 1). Refusers were older than participants (mean

920 British Journal of General Practice, December 2003
age 45.3 years versus 39.7 years; P<0.001) and more of them were diagnosed as somatisers by GPs (19.5% versus 12.6%; P<0.001).

Of the participating patients 82.3% had a symptom duration of less than 6 months and 32.1% scored positive on the screening questionnaire. Randomised groups only differed in two respects at the patient level. The inclusion rate was 81% in the intervention group and 72% in the control group, and men accounted for 41.3% of the intervention group and 35.0% of the control (P<0.001).

Effect of intervention on the GPs' classification

Dichotomisation of diagnoses into the categories 'physical' and 'non-physical' (somatisation) revealed that 14.2% of the main problems were classified as 'non-physical' by intervention GPs, compared with 10.8% for control GPs (diagnostic difference 3.5%, 95% confidence interval [CI] = -0.6 to 7.5, P = 0.094; adjusted for patient sex and clusters). Clustering was found to have a strong impact on the results as the estimated intra-cluster correlation coefficient was 0.027 (95% CI = 0.008 to 0.045) and clusters were large. The proportion of patients who were diagnosed as somatisers showed substantial variation between GPs, even within practices, with a mean value of 13.5% (range 3.3–33.9%) in the intervention group and 10.3% (range 2.5–21.9%) in the control group. This variation could not be accounted for by differences between GPs' patient populations assessed by scores on the screening questionnaire (for additional data see Supplementary Table 1 and Supplementary Figure 1).

Analyses of the original five-item questionnaire revealed a significant difference in the overall classification (P = 0.049, adjusted for patient sex and clusters). GPs in the intervention group classified twice as many patients with 'medically unexplained symptoms' as did control GPs (P = 0.007), and classified fewer with 'physical disease' (P = 0.085) (Table 2). Subgroup analyses showed no relationship between classification and GPs' sex, age, number of listed patients, participation in previous courses or supervision groups. However, the numbers in the subgroups were small.

The skewed patient inclusion necessitated separate analysis of those who refused to participate in the study. Like participants, refusers were more often classified as somatisers in the intervention group (22.1%) than in the control group (18.3%) (adjusted diagnostic difference 3.0; 95% CI = -6.8 to 12.7; P = 0.551).

Effect on classification compared with rating scales

In both groups GPs diagnosed more patients as somatisers if the screening questionnaire was positive (19.5%) than if it was negative (9.3%). The agreement between GP assessments and screening questionnaire results did not differ significantly between groups (Table 3).

For each GP we calculated sensitivity and specificity using the score from the screening questionnaire as the reference value. The sensitivity varied considerably between GPs, with a mean of 19.9% (SD = 11.4) in the intervention group and 16.6% (SD = 8.9) in the control group. The mean specificity was 89.9% (SD = 7.1) in the intervention group and 92.2% (SD = 5.1) in the control group.

Subgroup analyses indicated effect modification by GPs' sex and attendance at previous longer courses, but not by GPs' age, number of listed patients or previous attendance in supervision groups. The difference in agreement between the intervention and control groups was 3.01 for female GPs compared with -6.04 for male GPs (sex difference 9.05, 95% CI = 3.11 to 14.99; P = 0.003). For GPs who had previously attended courses, the difference was 1.70 compared with -6.39 for other GPs (difference = 8.09; 95% CI = 0.14 to 16.03; P = 0.046). However the results obtained from subgroup analyses must be interpreted with caution, as this study was not designed to investigate the effect of GP factors.

Discussion

Summary of main findings

A large and representative number of GPs in Vejle County participated in the study, which indicates that the results may be generalised to a similar setting. GPs diagnosed somatisation less frequently than has been observed in previous studies, but there was substantial variation between GPs. Our intervention did not significantly influence the overall classification of symptoms, but it did increase GPs' awareness of medically unexplained physical symptoms. The intervention failed to improve diagnostic accuracy assessed by the use of rating scales.

Strengths and limitations of this study

The study was conducted as a randomised controlled trial during routine practice, making even small effects valuable. The intervention addressed both diagnosis and treatment of the whole spectrum of somatisation, as recommended previously. To strengthen the study further, cluster randomisation was performed at practice level, thereby limiting contamination of the control group.

The effect of the intervention may have been overestimated, as intention-to-treat analyses could not be performed. However, the numbers of GPs leaving the study were identical in both groups indicating a non-differential dropout. Selection bias may also have contributed to the difference. Despite successful GP randomisation, the control group included fewer patients than the intervention group and patients who refused to participate in the study were more likely to be diagnosed as somatisers.

On the other hand, the effect of the intervention may also have been underestimated; first, by GPs being forced to make only one diagnosis despite being taught the complexity of problems, secondly, by the intervention targeting treatment more than diagnostic criteria for somatisation, and thirdly, by the fact that changing GPs' behaviour is a complex process which may take much longer than the limited time that was available in this study. A Hawthorne effect would have been present in both groups, and would tend to reduce the magnitude of differences. Furthermore, the number of enrolled GPs and patients did not reach the planned sample size and the cluster effect was found to be considerable. Both of these factors served to make the study underpowered.

The assessment of diagnostic accuracy poses several problems. Most of the patients in this study presented symptoms that had a duration of less than 6 months, so the
ICD-10 diagnosis for somatoform disorder could not be applied. Furthermore, diagnoses in general practice are made over time, whereas the assessment of accuracy is often cross-sectional. We chose rating scales for the measurement of accuracy. These are subject to a high degree of uncertainty, and do not always reveal whether a problem noted is related to the reason for the encounter. Finally, GPs’ diagnosis of somatisation has previously been shown to be only weakly associated with questionnaires based on psychiatric diagnostic classifications. However, these scales are currently the best available tools for rating somatisation by the use of questionnaires.

**Comparison of our findings with the literature**

Only 12.6% of the patients in our study were diagnosed as somatisers, compared with 20–30% of those in previous studies. This could be a result of our categorisation. Altering the dichotomy to ‘entirely physical disease’ and ‘others’ would yield figures close to those reported previously, but would not be consistent with our conception of somatisation. Another reason for the low diagnostic rates could be that participants represented a large proportion of GPs from Vejle County, including those without a special interest in psychiatric disorders. Finally, a diagnostic rate similar to that found in our study was reported by Weich et al, when they looked at the number of patients who presented emotional distress of relevance to their reason for consulting a GP.

Previous studies have shown that GP training may affect diagnosis of mental disorders and interviewing skills, but a recent study on depression failed to show any effect of education on recognition. In the same study, members of the Royal College of General Practitioners in the United Kingdom showed increased sensitivity in the diagnosis of depression. As was indicated in our study, this was at the cost of decreased specificity when compared with rating scales. The possible increase in sensitivity is consistent with previous data on somatisation in follow-up studies, and our study has added the certainty of a cluster randomised design.

**Implications for future research and clinical practice**

Training of GPs may increase their awareness of medically unexplained physical symptoms in relation to their patients’ reasons for presentation. To our knowledge, this has not been previously demonstrated in a randomised controlled trial. However, the accuracy of an increased number of diagnoses of medically unexplained symptoms is uncertain and requires further investigation. Assessment of GPs’ diagnoses depends on the establishment of a gold standard for somatisation that is applicable to general practice, and should be supplemented by evaluation of longitudinal diagnoses.

**References**


**Acknowledgements**

We would like to thank all of the patients and GPs in Vejle County who took part in this study. We are grateful to Ineta Sokolowski and Morten Frydenborg for extensive assistance with statistical analyses. Finally, we wish to acknowledge the support of The Quality Improvement Committee for General Practice in Vejle County (Q2), The Foundation for Medical Science in Vejle County, The Danish National Research Foundation for General Practice, The Regional Health Insurance in Vejle County, The General Practitioners’ Foundation for Education and Development and, Grants from the Foundations of Sara Kirstine Dalby Krabbe, Else Nicolaelsen and Dr K Rasmussen.

**Supplementary information**

Additional information accompanies this paper at \[http://www.rcgp.org.uk/rcgp/journal/supp/index.asp\]