Evaluating the effectiveness of 2 educational interventions in family practice

Alexander E.M. Borgiel,††† MD; J. Ivan Williams,*+‡§ PhD; David A. Davis,‡¶** MD; Earl V. Dunn,†+++ MD; Neil Hobbs,**+++ MB BCH; Brian Hutchison,+‡§§ MD; C. Ruth Wilson,+++ MD; Jamie Jensen†⁺⁺; Jennifer J.S. O’Neil,+++ BSc; Martin J. Bass,+++ MD, MSc

Abstract

Background: Structured feedback of information can produce change in physician behaviour. The objective of this study was to assess the effectiveness of 2 educational interventions for improving the quality of care provided by family physicians in Ontario: the Practice Assessment Report (PAR) and the Continuing Medical Education Plan (CMEP) with a follow-up visit by a mentor.

Methods: The study was a randomized controlled trial. Physicians in the control group received only the PAR, whereas those in the experimental group received the PAR, CMEP and mentor interventions. The participants were 56 family physicians and general practitioners (27 in the PAR group and 29 in the CMEP group) in southern Ontario who agreed to participate in the interventions and provide data. A total of 2395 patients randomly sampled from the practices returned questionnaires and consented to have their medical records abstracted. The outcome measures were global scores in 4 areas — quality of care, charting, prevention and overall use of medications — and patient ratings of satisfaction with care and preventive practices. The measures were applied at the beginning (phase 1) and end (phase 2) of the study.

Results: The mean global scores at the end of the study for the PAR group were 70.1% for quality of care, 84.7% for prevention, 77.7% for charting and 82.2% for overall use of medications. The corresponding scores for the CMEP group were 68.3%, 82.1%, 76.4% and 83.2%. In the patient satisfaction component, the personal care scores at phase 2 were 93.6% for the PAR group and 94.6% for the CMEP group. Examples of the scores for prevention for the PAR group were 98.3% for children’s current immunization, 96.6% for blood pressure measured within the previous 5 years, 79.4% for referral of women of the appropriate age for mammography within the previous 2 years, and 58.4% for discussion about alcohol use. The corresponding scores for the CMEP group were 95.8%, 97.6%, 77.6% and 64.6%. The changes in mean scores between phase 1 and phase 2 ranged from –1.9 to 2.3 points. There were no significant differences between the 2 groups in phase 1 or phase 2 scores or in change in scores. A total of 64.3% of the physicians rated the PAR as useful, 26.5% found the CMEP to be useful, and 41.0% considered the mentor strategy to be a useful form of continuing medical education. Although changes in practice related to the PAR, CMEP or mentor were reported by some physicians, they were not related to chart audit or patient scores.

Interpretation: Educational interventions based on quality-of-care assessments and directed to global improvements in quality of care did not result in improvements in the outcome measures. Educational interventions may have to be targeted to specific areas of the practice, with physicians being monitored and receiving ongoing feedback on their performance.
before 1998 physicians who had not completed a residency in family medicine could become candidates for certification by the College of Family Physicians of Canada (CFPC) by participating in the Practice Assessment Program. The CFPC entered the physicians’ practices, audited medical charts and asked patients to complete questionnaires. Explicit criteria were applied to the information from the chart audits and questionnaires, and the physicians received the Practice Assessment Report (PAR), which gave them a global assessment of quality of care without specific recommendations for changes in practice.

Structured feedback of information can produce change in physician behaviour.6 We designed a Continuing Medical Education Plan (CMEP) by structuring the information from the practice assessment to identify problems in practice and by directing physicians to continuing medical education (CME) activities that address the problems. The plan also called for mentors to visit the physicians in their practices, review the report and determine whether the structured feedback led to corrections of the problems without introducing new ones.6

Methods

Study design and recruitment

We designed a randomized controlled trial, conducted between April 1994 and June 1996, to compare the relative effect of the standard PAR and the experimental CMEP on the quality of family practice. Physicians in the control group received only the PAR, and those in the experimental group received the PAR, the CMEP and a visit from a mentor. Quality of family practice was assessed before and 1 year after the intervention.

To be eligible for the study, physicians had to have been in family practice for 5 years or more, to have been in the same practice for 3 years or more, to have graduated from medical school after 1958 and to work at least half-time in practice. Physicians who were specialists or full-time academics, were located north of French River, Ont., had been audited by the College of Physicians and Surgeons of Ontario or the CFPC within the previous 2 years or had participated in the Physician Review Program of the College of Physicians and Surgeons of Ontario were excluded.

We systematically sampled every fifth physician listed in the Ontario section of the 1988 Canadian Medical Directory and verified their listings in the 1990 and 1993 directories to ascertain length of time in family practice and in the same location. We stratified the sample so that one-quarter of the physicians came from Toronto.

Physicians who are female, are younger (under age 40 years) and have completed residency training in family medicine generally have higher scores for quality of care than do their counterparts.10-12 To balance the groups, we stratified physicians by sex, age and CFPC status into 8 permuted blocks. One of us (J.I.W.) prepared a separate list for random assignment for each block. When a physician agreed to participate, she or he was placed into the appropriate block, listed on the first open line and assigned to the PAR or CMEP group by one of the staff (J.J. or J.O.). The senior author verified that the physicians were listed in the order in which they were recruited.

Assessment of quality of care

The PAR and CMEP are based on 3 sources: patient questionnaires, chart audits and physician questionnaires.

Patient questionnaires

From appointment lists provided by the physicians, we randomly selected 30 sessions over 6 months and asked for the names and addresses of the first 5 patients seen in each session. Physicians were asked to include 1 patient per family and to exclude patients who were terminally ill or had died as well as the members of their families. Patients received a packet consisting of a covering letter signed by their physician, a questionnaire and a form to consent to having their chart abstracted. There were versions for children under 12 years of age, adolescents and adults. The materials were in either English or French. Patients were asked to return the questionnaire and the signed consent form to the research office, and we told them that their physician would not have access to their replies.

The questionnaire contained the following items: background information (age, sex and self-rating of health), patient care (personal care, brusqueness of care, access to care, patient–physician interaction and patient satisfaction with staff) and prevention (physician asked about smoking, asked about drinking history, measured blood pressure, discussed cholesterol testing, discussed influenza vaccination for older adults, discussed current immunization for children, performed breast examination, referred for mammography and obtained Papanicolaou smear).

Chart audits

A panel of family physicians, a college librarian and researchers modified the explicit criteria for 152 diagnoses,13 basing them on evidence when possible. We derived 4 global scores based on the proportion of criteria met for “quality of care,” “prevention,” “charting” and “overall use of medications.” The quality-of-care score was based on all applicable criteria for all diagnoses. The prevention score was based on how well the physician followed the recommendations14 for monitoring blood pressure, vaccinations, breast examinations, mammography, Papanicolaou smears for women of appropriate age, assessing cholesterol for appropriate ages and certain lifestyle behaviours. The charting score was based on legibility, organization and documentation of registration data, significant medical and family or social history, long-term medications, allergies and vaccinations. The score for overall use of medications rated how well physicians followed prescribing guidelines for indicator drugs: anticoagulants, antidepressants, NSAIDs, narcotics and tranquilizers.

The charts were assessed at the start of the study (phase 1) and about 12 months after the physicians completed the interventions (phase 2). We trained 8 people with working experience in family practice to abstract the records at the beginning of the study. At the end of phase 1 we replaced 4 abstractors who left the study and repeated the training experience for all of them. For phase 2 the abstractors applied the criteria to all patient visits during the preceding 12 months.

Physician questionnaires

Physicians completed questionnaires at the beginning and end.
of the study. The first questionnaire asked about practice characteristics, practice profile, provision of preventive manoeuvres, hospital practice, CME and self-rated performance with respect to clinical content elements and clinical practice elements. The second questionnaire repeated the self-ratings of performance but focused on self-evaluations of the educational interventions.

**Educational interventions**

The PAR did not provide any specific recommendations for improving practice, whereas the CMEP provided structured information for identifying relative strengths and weakness in the practice, including 3 areas of need for remedial action. Prevention was included as a need for all physicians, and the lowest scores on the clinical content elements and clinical practice elements were used to identify the 2 other needs.

Mentors, who were senior and respected physicians in the community, attended a workshop to practice the Practice Assessment Program and the educational interventions. They had 5 responsibilities: to answer queries about the practice assessment, to review scores in the PAR and CMEP, to suggest CME strategies for addressing the identified needs, to determine the physician’s learning style and to help the physician develop a learning plan. Mentors obtained commitments from the physician to undertake CME activities and sent a follow-up letter summarizing the visit and restating the expectations. Two or 3 months later the mentor called the physician and reviewed the commitments and subsequent activities.

At the end of the study, the mentors, abstractors and selected physicians from both study groups participated in focus groups, in which they debriefed the research team on their experiences and gave their assessments of the educational interventions and methods.

**Sample size estimation and data analysis**

We specified global quality of care as the key outcome variable for testing the effects of the educational interventions. For the CMEP and mentor interventions to be considered important, phase 2 scores for the CMEP group had to be 10 points higher than those for the PAR group. Based on results from the Practice Assessment Program, we estimated that the PAR group would have a mean score of 65 points. With a mean difference of 10 points, a standard deviation (SD) of 12 points, a 2-tailed test with an alpha of 0.05 and power set at 80%, the required sample size was 24 physicians per group.

We also tested for changes in scores over time. With samples of 24 physicians per group, we had sufficient power to test for group differences in score changes of 5 points, assuming the SD of change was 5 points or less. To ensure ample power for analysis, after allowing for attrition, we recruited 30 physicians per group.

We used unpaired t-tests of mean scores to test the effect of the CMEP relative to the PAR on the quality of practice, and paired t-tests to assess the changes within groups between phase 1 and phase 2.

**Results**

Of the 148 physicians sampled, 44 were ineligible for participation or could not be located. Of the 104 eligible physicians, 48 (46.2%) declined to participate, including 3 physicians who dropped out between the beginning and end of the study. Thus, 56 physicians completed the study, 27 in the PAR group and 29 in the CMEP group.

Participation rates varied significantly by sex, CFPC status and year of graduation. Women were more likely to participate than men (82.6% v. 45.7%, $\chi^2 = 9.83$, 1 degree of freedom [df], $p = 0.002$); 73.8% of certificated physicians participated, as compared with 40.0% of CFPC members and 40.4% of nonmembers ($\chi^2 = 11.30$, 2 df, $p = 0.004$); and physicians who graduated in 1970 or later were 3 times as likely to participate as physicians who graduated before 1970 (77.8% v. 22.3%, $\chi^2 = 8.18$, 1 df, $p = 0.004$). Table 1 shows the background characteristics of physicians assigned to the PAR and CMEP groups. The groups were balanced with respect to sex, CFPC status, practice location and year of graduation.

We sent 7957 questionnaires to patients of the 59 physicians who started the study, and the response rate was 57.7%. Fig. 1 shows the returns for the 56 physicians who completed the study by group; their patients returned 4401 questionnaires. As we included up to 50 charts per practice, abstractors audited 2787 charts. At the end of the study, we were able to match questionnaires and chart audits for 2395 patients from both phases of the study.

In both groups, 67% of the patients were female (Table 2). Overall, 14.6% were under 18 years of age, 58.2% were between 18 and 64 years, and 27.1% were 65 years or older; there were no differences by group. A total of 48.0% of the patients in the PAR group and 49.2% of those in the CMEP group reported their health as being very good or excellent.

The initial results from phase 1 showed a surprising distribution of scores for quality of care: 30% of the scores were above 90%, and another 25% were in the range of 80% to 89%. Either the physicians were very
good, or there were problems with the assessment methods. We randomly selected 25 charts from each practice and asked different abstractors to reabstract the phase 1 data. We compared the scores from the reabstracted data with the original scores for the 25 charts. The mean difference in paired scores for quality of care was 10.9 (SD 2.2), a statistically significant difference ($t = 4.88, p < 0.001$). The mean differences for prevention (1.4 [SD 0.9]), charting (3.3 [SD 1.5]) and use of medications (2.7 [SD 1.6]) were not significantly different. We discarded the original scores for phase 1 and replaced them with the reabstracted scores.

Table 3 shows the global scores for phase 1 and phase 2 and the change in scores by group. In the PAR group, the change in scores between phase 1 and phase 2 ranged from –0.4 for prevention to 2.3 for quality of care. In the CMEP group, the change in scores ranged from –1.9 for overall use of medications to 2.1 for quality of care.

None of the differences in global scores approached statistical significance, either within or between groups. When we used regression analysis to adjust for group differences in physician characteristics and for variations in patient age and sex across practices, the results remained essentially the same.

Patients expressed high levels of satisfaction with their physicians. The phase 2 mean scores for the PAR group and the CMEP group respectively were 93.6% and 94.6% for personal care, 85.5% and 84.7% for access to care, 85.1% and 85.3% for patient–physician interaction, and 93.7% and 89.1% for patient satisfaction with staff. On average, only 6.7% of the patients in the PAR group and 5.4% of the patients in the CMEP group characterized their physicians as brusque. The changes in patient satisfaction scores between phases 1 and 2 were within –3.3 to 1.4 points. None of the differences between and within groups approached statistical significance.

Generally, patients in both groups gave their physicians high ratings for prevention. In the PAR group in phase 2, 98.3% of children had current vaccinations, 96.6% of adults had had their blood pressure measured within the previous 5 years, 88.5% of older patients had been offered influenza vaccine, 79.1% of patients said their doctor had discussed smoking, 74.3% reported their physician had discussed cholesterol testing, and 58.4% of adults said their doctor had asked about their use of alcohol. Among the women, 74.6% had had a Papanicolaou smear within the previous 3 years, 79.4% had had a breast examination within the previous 2 years, and 76.2% of those aged 50 to 69 years had been referred for mammography within the previous 2 years.

In the CMEP group in phase 2, 95.8% of children had current vaccinations, 97.6% of adults had had their blood pressure measured within the previous 5 years, 83.5% of older patients had been offered influenza vaccine, 83.1% of patients said their doctor had discussed smoking, 73.3% reported their physician had discussed cholesterol testing, and 64.6% had been asked about their use of alcohol. Among the women, 79.9% had had a Papanicolaou smear within the previous 3 years, 77.4% had had a breast examination within the previous 2 years, and 77.6% of those of the appropriate age had been referred for mammography within the previous 2 years.

The change in scores for prevention between phase 1 and phase 2 ranged from –2.8 to 3.7 points across both groups (data not shown). Again, none of the differences was statistically significant.

In the follow-up questionnaire, 64.3% of the physicians rated at least 3 of the 4 parts of the PAR as being useful or very helpful. Physicians emphasized the importance of having results from patient questionnaires. The ratings did not vary by group. A total of 70.4% of the doctors in the PAR group and 82.8% of those in the CMEP group reported

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>PAR ($n = 1141$)</th>
<th>CMEP ($n = 1254$)</th>
<th>Total ($n = 2395$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>32.9</td>
<td>33.0</td>
<td>33.0</td>
</tr>
<tr>
<td>Female</td>
<td>67.1</td>
<td>67.0</td>
<td>67.0</td>
</tr>
<tr>
<td><strong>Age, yr</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 12</td>
<td>12.2</td>
<td>13.8</td>
<td>13.0</td>
</tr>
<tr>
<td>12–17</td>
<td>2.0</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>18–64</td>
<td>59.8</td>
<td>59.9</td>
<td>58.2</td>
</tr>
<tr>
<td>≥ 65</td>
<td>26.0</td>
<td>24.5</td>
<td>27.1</td>
</tr>
<tr>
<td><strong>Self-rating of health</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>14.4</td>
<td>15.8</td>
<td>15.1</td>
</tr>
<tr>
<td>Very good</td>
<td>33.6</td>
<td>33.4</td>
<td>33.5</td>
</tr>
<tr>
<td>Good</td>
<td>33.0</td>
<td>31.9</td>
<td>32.4</td>
</tr>
<tr>
<td>Fair</td>
<td>13.9</td>
<td>14.6</td>
<td>14.3</td>
</tr>
<tr>
<td>Poor</td>
<td>5.1</td>
<td>4.3</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Fig. 1: Numbers of patients who completed questionnaires and whose charts were abstracted in the practices of the 56 family physicians who completed the study. PAR = Practice Assessment Report; CMEP = Continuing Medical Education Plan.
that they had made changes in clinical practice as a result of the PAR; the difference was not statistically significant. Physicians in the CMEP group were more likely than those in the PAR group to report changes in CME activities (50.0% v. 20.8%; χ² = 4.74, p = 0.029).

Only 26.5% of the physicians in the CMEP group found most of the 5 parts of the CMEP to be useful or very useful. A total of 27.6% reported changes in CME activities in response to the CMEP, and 44.8% indicated changes in clinical practice. Most of the physicians (82.4%) rated the mentor visits as useful or very useful. Only 40.0% said the visits led to changes in clinical practice, 41.4% reported changes in CME activities, and 41.0% rated the mentor strategy as a useful form of CME. Self-reported changes in practice, however, were unrelated to changes in global scores.

**Interpretation**

Four major reviews of CME studies have been published since the start of our study.15-18 A wide variety of CME measures have been used to improve quality of care and outcomes, but, as Oxman and colleagues19 note, there are no “magic bullets.” Davis and associates20 reviewed 50 randomized controlled trials that involved mainly physicians, had complete follow-up assessments for at least 75% of the study subjects and provided objective assessments of either physician performance or health care outcomes. Of the 16 studies that focused on general clinical management, 12 showed positive changes in physician behaviours in response to interventions, but only some of the results achieved statistical significance; 2 demonstrated negative results; and 2 showed no effects, as did our study.

At the time we designed our study the quality assurance literature focused on management of specific diagnoses or interventions.19 Physicians’ records were audited to ascertain the degree to which their practices met evidence-based criteria for specific tracer or indicator conditions.20,21 In subsequent studies, academic detailing was introduced to modify prescribing practices related to overused drugs.22 Physicians’ practices have been assessed with respect to prevention practices, such as those recommended by the Canadian Task Force on the Periodic Health Examination.23-25 Patients have been taught to simulate specific conditions, and physicians have been rated on the care provided to them.24,25

Our study, in contrast, involved criteria and guidelines for the breadth and scope of family practice. With respect to the adequacy of the measures used in the study, the problems with the data abstraction call into question the reliability and validity of the scores, particularly those for global quality of care. We believe, however, that the retraining of abstractors and reabstraction of the data solved those problems, and we are confident that the methods we used for assessing the quality of care do work. Our methods have been reviewed by the Institute of Medicine,26 have been adopted for use in Boston,24 California and Manitoba.

It may be unrealistic to expect significant changes in practice following the receipt of a general report or structured feedback. The literature supports the use of trained physician-educators or opinion leaders for modifying practices. Mentors may be more effective in educating physicians about specific areas of practice in which the leader has expertise (e.g., prescribing of drugs) than when offering broad general consultations, as was done in our study. Interventions based on global assessments of practice are not a good means for identifying learning needs or targeting practice behaviours for change.

Martin Bass died while this project was in progress. Martin was a key participant in the study. We wish to acknowledge the leadership he provided to family medicine research in Canada.

This project was funded by a grant from the National Health Research and Development Programme of Health Canada.

Competing interests: None declared.

| Table 3: Global scores for quality of care and resource use by group and phase of study |
|-----------------|-----------------|-----------------|-----------------|
| Variable        | PAR group        | CMEP group       |
|                 | Phase 1          | Phase 2          |
|                 | No. of physicians| Mean score (and SD), % | No. of physicians| Mean score (and SD), % | Change |
| Quality of care | 27               | 67.8 (12.0)      | 27               | 70.1 (12.4)      | 2.3     |
| Prevention      | 27               | 85.1 (11.5)      | 27               | 84.7 (10.5)      | 0.4     |
| Charting        | 27               | 77.3 (12.5)      | 27               | 77.7 (12.4)      | 0.4     |
| Overall use of medication | 26    | 81.3 (12.8)      | 27               | 82.2 (11.9)      | 0.9     |
| Quality of care | 29               | 66.2 (11.3)      | 29               | 68.3 (10.6)      | 2.1     |
| Prevention      | 29               | 81.1 (9.7)       | 29               | 82.1 (8.7)       | 1.0     |
| Charting        | 29               | 75.4 (10.7)      | 29               | 76.4 (11.4)      | 1.0     |
| Overall use of medications | 29    | 85.1 (8.6)       | 29               | 83.2 (13.4)      | -1.9    |

Note: SD = standard deviation.
References


Reprint requests to: Dr. J. Ivan Williams, Institute for Clinical Evaluative Sciences, G-106, 2975 Bayview Ave., Toronto ON M4N 3M5; fax 416 480-6048