Effect of a PDA-assisted Evidence-based Medicine Course on Knowledge of Common Clinical Problems

Roland M. Grad, MD, CM, MSc; Yuejing Meng, MSc; Gillian Bartlett, PhD; Martin Dawes, MD; Pierre Pluye, MD, PhD; Miriam Boillat, MD; Goutham Rao, MD; Roger Thomas, MD, PhD

Background and Objectives: It is not yet known if personal digital assistant (PDA)-assisted evidence-based medicine (EBM) courses in postgraduate training enhance knowledge of common clinical problems. This study’s objective was to determine if PDA-assisted EBM training would improve clinical knowledge. Methods: In a controlled trial, intervention group residents received InfoRetriever on a PDA coupled with an EBM course integrated within clinical rotations in family medicine. The effect of the intervention and the rate of use of InfoRetriever on a written test of knowledge were evaluated after adjusting for baseline knowledge scores. The test measured knowledge of primary care management of hypertension and diabetes as well as estimation of disease probability. Results: There was no effect on first posttest knowledge scores of the intervention overall or of the rate with which participants had used InfoRetriever during the intervention. However, when intervention group residents retook the test with access to InfoRetriever while taking the knowledge test, scores increased 7.4% (+2.4 correct test questions). Access to InfoRetriever Clinical Prediction Rules on a PDA, however, had an unclear effect on residents’ ability to estimate disease probability. Conclusions: There was no effect of a PDA-assisted EBM course on knowledge test scores, although using the PDA during the test results in higher scores. It is unclear if using PDA Clinical Prediction Rules can improve residents’ estimates of disease probability.

(Fam Med 2005;37(10):734-40.)

Decision support tools and handheld computer (personal digital assistant [PDA]) summaries of evidence are increasingly used in residency training, since they enable the practice and learning of evidence-based medicine (EBM) in diverse clinical settings.\textsuperscript{1,2} It is often assumed that rapid access to PDA summaries of evidence will bring benefit to clinicians and their patients.

However, clinicians may have difficulty using PDA evidence-based search and appraisal techniques in a timely and efficient manner.\textsuperscript{3,4} Although teaching can have a positive influence on question formulation skills, use of Medline,\textsuperscript{5} and knowledge of epidemiology and biostatistics,\textsuperscript{6} it is not yet clear if EBM courses in residency training or PDA access to summaries of evidence have a positive effect on objectively measured knowledge of common clinical problems.\textsuperscript{7,8} This gap in our knowledge has been confirmed in a systematic review of the effects of residency teaching of EBM.\textsuperscript{9}

Sophisticated software tools available in PDAs now provide access to both clinical decision support systems (CDSS) and clinical information in databases that are largely or entirely text based. CDSSs require users to enter patient-specific data (Figure 1), and the system then returns information pertaining to risk levels or prognoses. Text-based databases, which we refer to as clinical information retrieval technology (CIRT) provide information not linked to specific patient data. Instead, CIRT includes databases of information about diseases, therapies, and interpretation of diagnostic tests, potentially applicable to deci-
sions about multiple patients.10 While the printed word predominates, CIRT may include images, sound, and movies, as well as multimedia.11 Patient-oriented Evidence That Matters (POEMs) in InfoRetriever are an example of CIRT.

In a randomized controlled trial using InfoRetriever on PDA, medical students in the intervention arm reported improved learning of EBM and increased confidence in clinical decision making. This study was limited, however, by self-reported measures of InfoRetriever use, since such measures are prone to recall bias.12,13 Further, this study did not separately assess the influence of the different databases searched by InfoRetriever. For example, the effect of using a CDSS might be different from that of text-based clinical information (eg, POEMs).14

Thus, additional research is needed to clarify the benefit, if any, of using PDA-based technologies, both CDSSs and CIRT, on trainees’ knowledge and their patient care. We coupled InfoRetriever on PDA with an EBM course in a cohort of first-year family medicine residents and assessed the effect of this intervention on objectively measured knowledge.

Methods

Our study had three stages: (1) development of an instrument to measure knowledge, (2) validation of this instrument, and (3) a before/after assessment of the effect of a PDA-assisted course in EBM. The instrument was developed by six family physicians and an endocrinologist.

Instrument Development and Validation (Knowledge Test)

In our setting, essential hypertension and diabetes are commonly encountered conditions whose management is well supported by evidence-based recommendations. In April 2003, we searched the literature for relevant and valid knowledge tests about these two conditions. For hypertension, no knowledge test with demonstrated validity was found. Therefore, 18 questions testing primary care knowledge of hypertension (diagnosis, treatment, and prognosis) were generated from multiple sources.15-20 For diabetes, 12 multiple-choice and true-false questions were taken from a validated “Family Physician Knowledge and Practice Survey.”21

Our knowledge test also assessed a physician’s ability to estimate the likelihood of streptococcal pharyngitis and coronary heart disease from information provided in a clinical case. To determine if use of PDA-based clinical prediction rules improved physician estimates of these probabilities, five “key feature” questions were written.22,23 Key feature questions were linked to clinical vignettes on sore throat and coronary heart disease. Sore throat vignettes required user entry of data on five clinical criteria into a validated clinical prediction rule.24 Vignettes on the 10-year risk of coronary heart disease (CHD) required user entry of data on eight variables into a clinical prediction rule derived from the Framingham study.25 The final knowledge test consisted of 35 questions and could be completed in 30 minutes or less. Test questions were specifically designed to have answers that could be found within InfoRetriever.

Reliability and validity of the knowledge test were assessed. We assessed validity by measuring the ability of the test to discriminate between three groups of physicians with different levels of knowledge (family medicine faculty versus first-year residents versus third-year medical students).26 We had more than 90% power to detect a 20% difference in mean score between these groups. Reliability was estimated by Cronbach’s coefficient alpha, calculated among test items at baseline. Endorsement rates and item response curves were examined for all multiple-choice questions in the baseline and the first posttest. Correct answers scored 1 point, wrong answers scored zero, and partial points were awarded for six questions (multiple true or false and short-answer open-ended questions).
Before and After Assessment

We conducted a controlled trial with first-year family medicine residents at McGill University. The trial had a parallel group design with before and after measurements. In June–July 2003, all 44 incoming residents based at four sites were asked to participate in a research study that offered a new PDA and training in software tools. In exchange for participation, PDAs were to be considered their personal property.

Consenting residents at sites A and B were allocated to an intervention group that received an EBM course and a PDA containing three software tools for EBM (InfoRetriever, versions 2003 or 2004), procedure tracking (Praxis), and drug prescribing (Lexidrugs).27,28 Residents at sites C and D were allocated to a control group and received the same tools, less the EBM course and InfoRetriever. Group allocation was made on pragmatic grounds, since the principal investigator runs an EBM course at sites A and B.28

The study protocol was reviewed and approved by the Faculty of Medicine Institutional Review Board of McGill University. Participants were unaware of our objectives prior to signing an informed consent form.

Intervention

The intervention incorporated InfoRetriever on PDA, which we had tested in the context of a pilot study.30,31 The InfoRetriever search engine allows simultaneous searching of seven databases: an electronic textbook (5-Minute Clinical Consult), the POEMs database, abstracts of Cochrane reviews, and guideline summaries, as well as CDSSs such as clinical decision and prediction rules, diagnostic test calculators, and history and physical exam calculators. Unique to commercially available software for primary care, InfoRetriever can automatically record data on information-seeking behavior derived from the users’ tap pattern. With participants’ consent, we used the InfoRetriever tracking function to document each opened item of information as an “information hit” in a log file on PDA. Participants transferred log file data from their PDA to our server upon synchronization. Log files provided specific data on information hits viewed by the resident, with each hit defined by a title and unique ID number, when the information was opened (date and time stamp), and what search strategy was used.

Prior to the EBM course, intervention group residents received 3 hours of software training divided into two sessions, one of which was devoted to using InfoRetriever. The second training session fell at the beginning of a 2-month family medicine block rotation during which participants attended the EBM course (fall 2003 or spring 2004). Further InfoRetriever training was offered during the eight weekly course seminars. Thus, the intervention group received InfoRetriever training in a reiterative fashion. Within the EBM course, intervention-group residents learned about clinical question formulation, quantitative understanding, and integrating new clinical knowledge in patient care.

Control group residents received their usual curriculum. In addition, they were trained to use Praxis and Lexidrugs. They did not, however, attend an EBM course, and they did not receive InfoRetriever.

Administration of the Knowledge Test

We administered the knowledge test to all participants just before the first EBM course (fall 2003) and again 8 months later. We changed the order of test items at the posttest to minimize any learning effect. Residents and students were directly observed during testing, at which time books, PDA software, or calculators were not permitted. Immediately after completing the first posttest, intervention-group residents were asked to “Retake the test using InfoRetriever to try to improve your score,” where answers to all test questions could be found. At this second posttest, we determined if participants used two InfoRetriever clinical prediction rules on their PDA: the McIsaac Sore Throat Score and a coronary risk assessment tool based on Framingham data. We determined if access to InfoRetriever was associated with improved test scores and correct risk estimates of streptococcal pharyngitis and CHD. We defined “correct” as an estimated risk ± 5% of the absolute value of risk provided by the clinical prediction rule.

Data Analysis

Using unpaired Student’s t tests in SPSS (version 12 for Windows), we compared mean scores between groups at baseline and at the first posttest. The difference in mean scores between intervention and control groups on the first posttest was expressed as the absolute score difference.

Using linear regression modeling in SAS (version 8.0 for Windows), we estimated the effect of the intervention on knowledge as measured by the score on the first posttest while controlling for scores obtained in the baseline test. In a second regression model, we estimated the influence of PDA access on second posttest scores while controlling for scores obtained in the baseline test. In a final model, we estimated the effect of the rate of InfoRetriever information hits on knowledge improvement as measured by first posttest scores while controlling for baseline score differences. Rate of information hits was determined for each participant from PDA log files and defined as the number of uses divided by the number of study days. Any days of lost data were subtracted from the number of study days. The time window to define use began on the day after the first training session until the day before the posttests. In further subanalysis based on the results of five key feature questions for residents in the interven-
tion group only, a paired $t$ test was used to compare the number of correct risk estimates on the first and second posttests.

**Results**

**Participants**

A total of 37 first-year residents consented to participate. These participants included 20 of 23 eligible residents at sites A and B (87%) and 17 of 21 eligible residents at sites C and D (81%) (Table 1).

<table>
<thead>
<tr>
<th>Group</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>12 (60.0)</td>
<td>8 (47.1)</td>
</tr>
<tr>
<td>Female</td>
<td>8 (40.0)</td>
<td>9 (52.9)</td>
</tr>
<tr>
<td>Age &lt;26</td>
<td>7 (35.0)</td>
<td>2 (11.8)</td>
</tr>
<tr>
<td>26–30</td>
<td>10 (50.0)</td>
<td>8 (47.1)</td>
</tr>
<tr>
<td>31–40</td>
<td>2 (10.0)</td>
<td>4 (23.6)</td>
</tr>
<tr>
<td>&gt;40</td>
<td>1 (5.0)</td>
<td>2 (11.8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>1 (5.9)</td>
</tr>
<tr>
<td>North American medical graduate</td>
<td>17 (85.0)</td>
<td>13 (76.5)</td>
</tr>
</tbody>
</table>

Table 1

Demographic Characteristics of Participating First-year Residents (n=37)

<table>
<thead>
<tr>
<th>Test</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline test score (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% correct</td>
<td>18.2 (17.0–19.3)</td>
<td>15.9 (14.7–17.2)</td>
</tr>
<tr>
<td>n=20</td>
<td>56.9%</td>
<td>49.7%</td>
</tr>
<tr>
<td>First posttest score (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% correct</td>
<td>20.5 (19.1–21.9)</td>
<td>18.8 (17.1–20.5)</td>
</tr>
<tr>
<td>n=17</td>
<td>64.1%</td>
<td>58.8%</td>
</tr>
<tr>
<td>Second posttest score (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% correct</td>
<td>22.8 (21.3–24.3)</td>
<td>20.8 (19.0–22.5)</td>
</tr>
<tr>
<td>n=12</td>
<td>71.3%</td>
<td>58.8%</td>
</tr>
</tbody>
</table>

Table 3

Comparison of Knowledge Test Scores Between Intervention and Control Groups at Different Test Times

**Test Validation and Discrimination**

The baseline test was taken by 105 physicians (45 faculty volunteers, 37 residents, and 24 medical students). The instrument distinguished between these groups (Table 2). In our analyses, we excluded three test items, two of which had endorsement rates of 99% and one of which did not have a clearly correct answer. Thus, the maximum possible score was 32, while the minimum possible score was 0. In this reduced set of test items, Cronbach’s alpha was 0.504.

**Use**

We documented 5,160 InfoRetriever hits by the intervention group during the study period. Not unexpectedly, use was highest during EBM course months. The mean frequency of InfoRetriever use was 0.98 hits per day (95% CI 0.78–1.18), with substantial variation in use between residents (range 0.4 to 1.5 hits per day). The three most used databases were the guideline summaries, an electronic textbook (5-Minute Clinical Consult), and POEMs.

Thirteen of 20 intervention group residents (65%) experienced some loss of usage data due to “hard resets” from depleted batteries and hardware failure. On average, these technical problems resulted in 28.5 days of usage data lost per participant over 208 days of usage tracking (13.7%). In addition, an error in the InfoRetriever tracking function affected eight participants for an average of 55 days. This error failed to track one specific type of search strategy for clinical prediction rules and was fixed prior to the posttest.
**Before and After Assessment**

For academic or personal reasons, three residents did not complete the intervention, and one resident withdrew from the control group. Therefore, the first posttest was taken by 33 residents. Mean scores in both intervention and control groups increased from baseline to first posttest (Table 3). On the first posttest, there was a 5.3% absolute difference in mean test score in favor of the intervention group (absolute score difference of 1.7 correct test questions). Within groups, the percentage improvements in knowledge were 7.2% in the intervention group compared with 9.1% in the control group. A comparison of change scores also showed improvement in both groups (intervention group mean change score = +2.0 correct test questions (95% CI = 0.3–3.6) versus control group mean change score = +2.9 (95% CI =1.5-4.3)).

When controlling for scores obtained on the baseline test, however, there was no effect of the intervention on first posttest scores. In addition, there was no significant relationship between the amount of InfoRetriever use and first posttest scores. However, we found a 7.4% increase in test score (+2.4 correct test questions, \( P = .046 \)) among intervention group residents who completed the second posttest with access to InfoRetriever on PDA. We then examined in further detail the estimates of disease probability provided by residents who took the second posttest with access to InfoRetriever (n=12).

With regard to estimating CHD risk and streptococcal pharyngitis, individual probability estimates provided by all participants for each clinical vignette spanned a very wide range at baseline, with both groups overestimating the absolute risk. The percentage of residents correctly estimating the risk of streptococcal pharyngitis as low improved in both groups from baseline to first posttest (Table 4). At the second posttest, the percentage also increased—from 65% to 92%, but as this measure included only the subgroup of responding residents (ie, smaller denominator), the absolute number of residents providing a correct response was unchanged (Table 5). The number of correct probability estimates for other items on the second posttest also increased, but again, decreasing denominators were involved. Interestingly, two of 12 intervention-group residents who responded to the second posttest did not use the relevant InfoRetriever clinical prediction rules, even though they had them on hand.

**Discussion**

There are several implications of this study for medical education. First, there was no significant effect of the intervention on first posttest knowledge scores. Therefore, the idea that a PDA-assisted EBM course would increase resident knowledge of common clinical problems is not supported. Further, it is unclear whether our intervention enhanced residents’ ability to correctly estimate the probability of streptococcal pharyngitis and CHD risk. This issue requires further study, because the increased percentage of correct responses may have been an artifact related to a decrease in the

<table>
<thead>
<tr>
<th>Vignette (Clinical Prediction Rule Probability)</th>
<th>Group</th>
<th>Mean Risk Estimates at Baseline (SD)</th>
<th># of Residents With Correct Estimate at Baseline</th>
<th># of Residents With Correct Estimate at First Posttest</th>
<th># of Residents With Correct Estimate at Second Posttest (With Access to Clinical Prediction Rules)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Low (1%) probability of streptococcal pharyngitis</td>
<td>Intervention</td>
<td>16.5% (14.1) 15.0% (14.2)</td>
<td>35% (7/20) 12% (2/17)</td>
<td>65% (11/17) 56% (9/16)</td>
<td>92% (11/12)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>15.0% (14.2) 14.1% (14.1)</td>
<td>12% (2/17) 12% (2/17)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. High (35%) probability of streptococcal pharyngitis</td>
<td>Intervention</td>
<td>76.4% (12.4) 76.1% (13.3)</td>
<td>0% (0/20) 0% (0/17)</td>
<td>12% (2/17) 13% (2/16)</td>
<td>67% (8/12)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>76.1% (13.3) 76.2% (13.4)</td>
<td>13% (2/16) 13% (2/16)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. CHD risk: non-smoker (17%)</td>
<td>Intervention</td>
<td>26.7% (19.5) 17.8% (10.8)</td>
<td>35% (7/20) 12% (2/17)</td>
<td>35% (6/17) 31% (5/16)</td>
<td>50% (6/12)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>19.5% (19.5) 10.8% (10.8)</td>
<td>29% (5/17) 29% (5/17)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4. CHD risk: smoker (25%)</td>
<td>Intervention</td>
<td>46.3% (27.5) 39.4% (23.8)</td>
<td>20% (4/20) 12% (2/17)</td>
<td>29% (5/17) 31% (5/16)</td>
<td>58% (7/12)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>27.5% (27.5) 23.8% (23.8)</td>
<td>25% (4/16) 25% (4/16)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5. CHD risk: diabetic (22%)</td>
<td>Intervention</td>
<td>55.4% (25.6) 42.7% (25.2)</td>
<td>10% (2/20) 29% (5/17)</td>
<td>41% (7/17) 25% (4/16)</td>
<td>83% (10/12)</td>
</tr>
</tbody>
</table>

CHD—coronary heart disease
PDA—personal digital assistant
of disease probability estimates to estimate disease probability. On the other hand, residents could not increase their scores by correctly answering test questions that required retrieval of text-intensive information. This finding challenges the assumption that simply providing PDA software to residents will enhance their knowledge of common problems seen in primary care. Residents face several barriers when trying to answer clinical questions.

In trying to find answers to test questions by retrieving text-intensive information on PDA, it may be that residents simply missed the answer located within the text of a POEM or a guideline. Qualitative analysis of interviews with participants revealed a problem of “missed” information, as illustrated by the following example. In the context of his daily work, a resident saw a patient with an unusual skin rash. He searched InfoRetriever on PDA to determine if the distribution of the rash was characteristic of molluscum contagiosum and told us that InfoRetriever provided no information to address his clinical question. However, our InfoRetriever search revealed the answer he missed, which we located in the first sentence of the 5-Minute Clinical Consult overview.

Our results do, however, support current efforts related to training in the use of clinical prediction rules. In other work, providing computer-generated coronary risk profiles to doctors in primary care was associated with greater improvements in serum lipid profiles and coronary risk among high-risk patients.34 Our finding that residents overestimate disease probability are consistent with two other studies in which community-based physicians overestimated the absolute risk of streptococcal pharyngitis and coronary disease in individual patients.35,36 Therefore, teachers should integrate handheld clinical prediction rules in family medicine residency training. In contrast, further research is needed to contribute to our knowledge of the benefits of educational interventions using CIRT.

Acknowledgments: This study was funded by the Canadian Institutes of Health Research, grant KTS-62367. Dr Grad is supported by the McGill University Department of Family Medicine and the Herzl Family Practice Center. Dr Pluye is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research.

We thank Armand Aalamian, Randi Morris, Maureen Doyle, and Suzanne Bouchard for contributing to the training of participants and Sara Meltzer, David Topps, and Gene Bailey for contributing to knowledge test development.

Corresponding Author: Address correspondence to Dr Grad, McGill University, Department of Family Medicine, Herzl Family Practice Centre, 3755 Cote-St-Catherine Road, Room E-0010, Montreal H3T 1E2, Quebec, Canada. 514-340-8222 # 5851. Fax: 514-340-8300. Roland. Grad@mcgill.ca.

References