Teaching Critical Appraisal: A Pilot Randomized Controlled Outcomes Trial in Undergraduate Osteopathic Medical Education

Paul M. Krueger, DO

Context: Critical appraisal is an important skill for medical students. A proposed curriculum may be an effective teaching tool.

Objective: To determine whether the teaching of critical appraisal can be successfully introduced into an osteopathic clinical clerkship in obstetrics and gynecology.

Design: Osteopathic medical students (N=77) were assigned by lottery to one of eight rotation groups during their clinical clerkship in obstetrics and gynecology. Four of these rotation groups received instruction in critical appraisal (study group; received evidence-based medicine [EBM] curriculum; n=38); the other four rotation groups did not (control group; received non-EBM; n=39). The ability of the study EBM group to critically analyze the literature was compared with that of the control (non-EBM) group on the basis of results of a multiple-choice examination.

Setting: The University of Medicine and Dentistry of New Jersey–School of Osteopathic Medicine clinical clerkship in obstetrics and gynecology.

Results: The median scores for critical analysis were 41 for the control group and 64 for the study group. This difference was statistically significant (P<.001).

Conclusion: The teaching of critical appraisal can be successfully introduced into a clerkship in obstetrics and gynecology.

J Am Osteopath Assoc. 2006;106:658–662

In the practice of evidence-based medicine (EBM), physicians conscientiously, explicitly, and judiciously use the current best evidence in making decisions about the care of individual patients. The inference is that physicians practicing EBM will have excellent skills in critical appraisal, the ability to dissect medical literature, analyze strengths and weaknesses of studies, and adjust their approach to patient care accordingly.

To facilitate the practice of EBM, the US Preventive Services Task Force rates studies by the quality of evidence that they provide. The highest level of evidence in this system, and EBM in general, is the randomized controlled trial (RCT). Few trials studying the introduction of critical appraisal exist; fewer still have examined critical appraisal in an undergraduate medical curriculum. Few used a study design that includes randomization and control; those that did have serious methodologic flaws.

In their 1998 review of the literature, Norman and Shannon identified 10 studies of the impact of teaching critical appraisal skills to medical students vs residents. They eliminated three of these studies because they were methodologically unsound, simply reporting the process of teaching critical analysis or using some form of “happiness index” (eg, a student’s subjective evaluation that the study was worthwhile). None of the studies used random assignment of participants into study and control groups. Norman and Shannon concluded that gains in knowledge at the undergraduate level consistently improved as a result of teaching critical appraisal skills, whereas only small changes in knowledge occurred at the residency level.

In reviewing critical appraisal on a graduate medical education level, Green found only four studies in a search of databases covering 1973 through 1998 that met minimal methodologic standards (ie, a pretest-posttest controlled trial).

In a more recent study, Taylor et al investigated a half-day program for practicing physicians in the United Kingdom. The study was designed as an RCT. It demonstrated improved overall knowledge scores and the ability of the physicians to critically appraise results of a systematic review at 6 months. It did not demonstrate differences in attitudes toward EBM, evidence-seeking behavior, perceived confidence, or other critical analysis skills.

The present pilot study shares a study design that is similar to that of Taylor et al, the RCT, avoiding the methodologic limitations of previous studies. Its purpose is to determine if the proposed curriculum is an effective way to increase medical students’ knowledge of critical appraisal.
Methods

Experimental Design

The present study was designed as an RCT to determine the effectiveness of introducing an EBM curriculum to osteopathic medical students during their 6-week clinical clerkship in obstetrics and gynecology for teaching critical appraisal. Students were randomly assigned to study (EBM) and control (non-EBM) groups. The EBM curriculum, taught entirely by the author, was given as part of a daily series of group workshops and lectures and took 6 hours of faculty time for each EBM study group to complete. The non-EBM groups received other lectures not related to critical appraisal.

At the end of the 6-week rotation, all students—both those in the EBM and those in the non-EBM groups—were given a multiple-choice examination covering core content in obstetrics, gynecology, and women’s health. This test counted for 40% of the students’ grade for the clerkship. The students also took a structured oral examination (20% of the clerkship grade). Clinical evaluation forms using a scale of 0 to 40 constituted the remaining 40% of the grade. In this way, each student received a numerical grade for the clerkship.

Following the first examination, a second multiple-choice examination was given to all students in both groups. It included articles and excerpts from articles to be reviewed by the students. Questions on critical analysis were devised relevant to each article or abstract. The test included generalized questions about EBM. In addition, students were expected to determine study design and know the strengths and weaknesses of each design. They were asked to evaluate bias and error in each study, and they were expected to rate the level of evidence and make evidence-based patient care recommendations.

The examination on critical analysis did not count toward the students’ grade in obstetrics and gynecology. To minimize communication between groups, students were notified following the examination that the critical analysis part of the examination was not reflected in their clerkship grade. They were told that they were part of an RCT and asked not to communicate this information to other students. The University of Medicine and Dentistry of New Jersey–School of Osteopathic Medicine (UMDNJ-SOM) institutional review board reviewed the process and decided that informed consent was not needed.

The examination was independently reviewed. Two reviewers with expertise in medical education who were blinded to the study-arm assignment, established face validity. They agreed that the examination assessed critical analysis and the curriculum taught the body of knowledge students would need to demonstrate expertise in critical analysis.

The examination was analyzed. Each question was placed in one of three categories. The first category dealt with study design and the forms of bias that were most prominent in each design. The second category dealt with statistics, and the third category dealt with applying the conclusions of the articles to clinical practice and making appropriate evidence-based patient care recommendations. The difference in the percentage of students correctly answering the questions between students who had received the EBM curriculum and those who had not were compared for each category of question. The mean scores were compared using a 2-tailed t test.

The UMDNJ-SOM does not calculate grade point average (GPA) in any year; therefore, the EBM and the non-EBM groups were compared for undergraduate admission criteria, including scores on the Medical College Admission Test (MCAT), undergraduate GPA, and percentage in each group from schools rated most competitive or higher by Barron’s Compact Guide to Colleges.6 Performance of the group receiving the EBM curriculum and the group receiving the non-EBM curriculum according to sex was also compared, and Barron’s Compact Guide to Colleges rating of students’ schools using the Pearson’s product moment χ² test was evaluated. The GPAs and the MCAT and examination scores were analyzed using the t test for equality of means.

The study did not include a crossover arm.

Study Participants

Students at UMDNJ-SOM rotate through a 6-week clinical clerkship in obstetrics and gynecology in the third year of their undergraduate curriculum. For 3 weeks, the students are assigned to in-hospital rotations in labor and delivery (2 weeks) and gynecologic surgery (1 week) at one of three sites. The 3-week ambulatory portion of the service takes place at an inner-city hospital clinic (1 week) and a suburban primary care facility for indigent patients with a large women’s health section (2 weeks). Much of the teaching takes place at the latter facility.

The students are assigned by lottery to eight groups, with each group comprising 8 to 10 students. These groups rotate together through the various clinical clerkships.

The rotating clinical clerkship groups for the 1998–1999 academic year were randomly assigned to a study (EBM) or a control (non-EBM) group. Eight groups of students rotated through the obstetrics and gynecology clerkship that year. The non-EBM group received the standard curriculum in obstetrics, gynecology, and women’s health as it had been taught for the previous 2 years. Although this curriculum addressed some issues in EBM and critical appraisal, there was no formal EBM curriculum. The EBM group received the usual curriculum plus a curriculum of critical appraisal that was integrated into the 6-week clerkship.

Evidence-Based Medicine Curriculum

The EBM curriculum that was used in this pilot study is an adaptation of a teaching template previously reported by Grimes7 and included the following components:

Lectures—A formal lecture based on EBM was given to each group. The lecturer defined and explained the importance of EBM.
Discussions—Each study group had two separate and distinct small-group discussions on critical appraisal that covered how to read an article, understanding potential sources of bias, and how to analyze a study and make clinical decisions based on that analysis.

Reading Expectation—Students were required to read the American College of Obstetricians and Gynecologists Practice Pattern titled Reading the Medical Literature: Applying Evidence to Practice.8

Journal Club on the Relationship of Pelvic Inflammatory Disease and the Intrauterine Device—The students were given an assignment to read several classic articles.9–12 They discussed the individual study designs and how the biases were introduced into these studies and influenced the outcome and conclusions. Students were then given an assignment to read an article reviewing these biases.13

Use of the Cochrane Library—Students received tutelage in the use of the Cochrane Library, a CD-ROM–based collection of meta-analyses.14

Evidence-Based Assignment—The students were presented with a clinical case scenario and asked to develop and review one of four pertinent clinical questions (Figure). They were expected to search articles, thoroughly review the literature, and critically appraise the articles that they found that answered the clinical questions.

Results
Follow-up was available for 100% of the students. All 77 students completed the clerkship and took the examination. Seventy-seven students completed the clerkship. Thirty-eight of these students were assigned to the EBM group and received the curriculum in critical analysis; 39 students were in the non-EBM group. One student in the EBM group failed the clerkship and had to repeat the rotation. His first (failing) scores were used.

Sex and admission demographics of the students are shown in Table 1. No significant differences existed between the EBM and non-EBM groups when evaluated for gender, MCAT score, GPA, and percentage of students from the undergraduate schools rated most competitive by Barron’s Compact Guide to Colleges.6 Randomization eliminated such factors as students’ aptitude and demographics. In addition, it used an acceptable evaluation tool—a multiple-choice examination.

Overall results on the multiple-choice examination appear in Table 2. The mean scores for the examination in critical analysis were 62% for the EBM group and 41% for the non-EBM group. This difference is statistically significant (t75, P<.001). The EBM group had a median score of 81 for the clerkship grade.

### Table 1
Characteristics of Study and Control Groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study Group (n=38)</th>
<th>Control Group (n=39)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Male</td>
<td>15</td>
<td>21</td>
<td>.29</td>
</tr>
<tr>
<td>□ Female</td>
<td>23</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Undergraduate GPA</td>
<td>3.55</td>
<td>3.53</td>
<td>.73</td>
</tr>
<tr>
<td>MCAT score</td>
<td>8.19</td>
<td>8.63</td>
<td>.151</td>
</tr>
<tr>
<td>From most competitive undergraduate colleges, %</td>
<td>52.6</td>
<td>51.3</td>
<td>.9</td>
</tr>
</tbody>
</table>

Abbreviations: GPA indicates grade point average; MCAT, Medical College Admission Test.

### Table 2
Critical Appraisal Examination Results and Clerkship Grade: Study Group vs Control Group*

<table>
<thead>
<tr>
<th>Grading Tool</th>
<th>Study Group (n=38)</th>
<th>Control Group (n=39)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical appraisal examination, %</td>
<td>62</td>
<td>41</td>
<td>.001</td>
</tr>
<tr>
<td>Clerkship grade</td>
<td>81</td>
<td>82</td>
<td>.547</td>
</tr>
</tbody>
</table>

* Study group received curriculum in evidence-based medicine; control group received standard curriculum.

---

**Case Presentation**
A 27-year-old, gravida 4, para 1, aborta 3 presents with vaginal spotting at 8 weeks of confirmed pregnancy. She is concerned because of her three previous spontaneous abortions (first, second, and fourth pregnancies). During her last pregnancy, she had a blood test and was told she had a “low progesterone.”

**Clinical Questions**

☐ **Etiology**: What is the significance of a low serum progesterone level in patients with spontaneous abortion?

☐ **Diagnosis**: What tests are appropriate to manage this patient?

☐ **Prognosis**: What is the chance of a patient with three consecutive spontaneous abortions having another one?

☐ **Therapy**: Is there an effective intervention?

---

**Figure. Evidence-based case study.**
and the non-EBM group achieved a median score of 82, a difference that was not statistically significant \((P=.547)\).

Table 3 shows the average percentage of correct answers by question category. Students receiving the EBM curriculum had a higher percentage of correct responses for questions that were categorized as identifying study design, statistics, and application to clinical practice. However, none of these differences was statistically significant, probably because of the small number of questions in each category.

### Comment

The results of the pilot study described here show that critical analysis can be taught. Students receiving the EBM curriculum clearly performed better on critical analysis than those who did not. This improved performance is not surprising; medical students are trained to reproduce information on multiple-choice examinations. This difference cannot be explained by the aptitude of the students. There was no statistically significant difference between the study and the control groups when overall grade for the clerkship was analyzed.

This study avoids many of the problems of previous published reports. The study design assures that biases are minimized. The results suggest that selection bias was effectively eliminated because the scores on the clerkship grade were the same in both groups. Observational bias was also minimized. The author taught all students in all workshops. Each student took the same examination.

Confounding biases were carefully considered. The UMDNJ-SOM does not calculate GPA or class rank; however, both groups were similar for preadmission MCAT scores and GPA. The ratings for their undergraduate schools were nearly identical. The similar score of the two groups for their overall clerkship grade suggests that the aptitude of the students was not different. Students’ sex was also considered because this author has previously reported that female medical students have superior performance during a clerkship in obstetrics and gynecology.\(^{15}\)

The current study also used an objective evaluation tool, a multiple-choice examination, rather than student polls or self-assessment (the so-called happiness index\(^3\)). Opinions differ as to whether this multiple-choice examination is best given as a pretest and posttest or only as a final examination.\(^{3,16}\) Because of the potential that a pretest would lead to information bias, with students in the control group anticipating the posttest and feeling that the posttest score would count toward their clerkship grade, the final examination format was chosen.

The analysis of the question responses by question category provides interesting information about the strengths and weaknesses of the curriculum. The curriculum was most effective at teaching clinical application, an important result because the ultimate purpose of the curriculum is to teach students how to care for patients. However, the only difference in students’ scores to reach statistical significance was on questions related to study design and bias. It is likely that this difference reflects the small number of questions in each subgroup leading to type II error. It was gratifying that the students did well on the statistical portion of the examination, suggesting that this curriculum is an effective way to teach a traditionally difficult topic.

The analysis of question responses by question category revealed the greatest variation between study subjects and control subjects in answers to questions that involved interpretation of statements quoted verbatim from articles. This difference suggests that the curriculum is meeting its goal—teaching critical appraisal. It also may reflect the fact that students are exposed to varying amounts of statistics and research design during their undergraduate and early osteopathic medical school training.

The analysis also showed that the difference between subjects who received the EBM curriculum and subjects who received the non-EBM curriculum was least marked on questions asking students to identify study design, another traditionally difficult topic that is not taught anywhere else in the osteopathic medical school curriculum at UMDNJ-SOM. It is likely that curriculum modification and, more important, repetition and reinforcement over the learning continuum will be needed to adequately teach study design.

However, the low overall score achieved by the students in the EBM group is disappointing. Several explanations are possible. The first is that the information was not taught as effectively early in the year as it was later in the year, thus lowering scores. However, the first group that received the EBM curriculum had a score of 64 and the last group that received the EBM curriculum had a score of 60. Therefore, it is unlikely that this explanation is correct.

Another possibility is that the test is too difficult. Therefore, the difficulty factor (ie, the proportion of examinees who answer each question correctly\(^{17}\)) was calculated. When the results of the study group were analyzed, 27% of the questions

### Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study Group</th>
<th>Control Group</th>
<th>(P) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying study design</td>
<td>0.635</td>
<td>0.511</td>
<td>.40</td>
</tr>
<tr>
<td>Statistics</td>
<td>0.715</td>
<td>0.385</td>
<td>.65</td>
</tr>
<tr>
<td>Clinical application</td>
<td>0.55</td>
<td>0.3</td>
<td>.88</td>
</tr>
</tbody>
</table>

Note: Study group received curriculum in evidence-based medicine; control group received standard curriculum.
had a difficulty factor of less than 0.5. However, most of the more difficult category of questions differentiated the upper quartile of students from the lower quartile. Thus, it appears that the test was indeed difficult.

Another likely explanation is that the students did not prepare for the critical analysis examination as they did for their clerkship examination. These students were not told that the critical analysis information would be part of their clerkship examination. They were told that the examination would be a comprehensive review of obstetrics, gynecology, and women’s health.

Future investigations into teaching EBM and critical analysis should focus on changing student and physician behavior. The results of McCluskey and Lovarini’s study with occupational therapists were disappointing. Although knowledge increased, changes in behavior such as frequency of searching and appraisal activities were small. The results are similar to those of the study of Taylor et al5 with physicians.

Conclusion
Medical students learn what they are taught. The analysis by question category shows that the EBM curriculum teaches study design and statistics but, more important, it shows that when students receive an EBM curriculum, they can interpret the medical literature.

Other questions need to be answered in the future. The curriculum described here was designed as an introduction only; it is not clear whether it is best taught in the second year of osteopathic medical school or during the clerkship years. Reinforcement in teaching study design is clearly needed. In addition, measures to apply this structure to other specialties need to be explored. The curriculum should be expanded to include other necessary parts of the EBM movement, such as decision analysis, critically appraised topics, and use of other EBM resources. In the meantime, educators can feel comfortable that an EBM curriculum will allow effective teaching of critical appraisal.

Acknowledgment
Sherry C. Pomerantz, PhD, an adjunct assistant professor in the Department of Internal Medicine at the UMDNJ-SOM in Stratford, contributed to the statistical analysis.

References


