

Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal

Geoffrey R. Norman, PhD; Susan I. Shannon, PhD

Abstract

Objective: To examine the evidence that the teaching of critical appraisal (evidence-based medicine) skills to undergraduate medical students or residents will result in significant gains in knowledge and increased use of the literature in clinical decision-making.

Data sources: Articles published from 1966 to 1995, retrieved through a MEDLINE search supplemented by manual searches; review of bibliographies maintained by individuals involved in teaching critical appraisal skills; and a previous methodological review.

Study selection: Articles were selected if the study involved some form of control group, although strict randomization was not required, and a measure of performance followed the intervention. Articles were excluded if they simply reported the process of teaching critical appraisal skills or used some form of "happiness index."

Data synthesis: There were 10 studies of the impact of teaching critical appraisal skills, 6 involving medical students and 4 involving residents. Results from 3 of the studies were nearly uninterpretable and thus were excluded; the remaining 7 were methodologically acceptable. Analysis showed that interventions implemented in undergraduate programs resulted in significant gains in knowledge, as assessed by a written test (mean gain 17.0%; standard deviation [SD] 4.0%). Conversely, studies at the residency level consistently showed a small change in knowledge (mean gain 1.3%; SD 1.7%). Two studies that examined residents' use of the literature were unable to demonstrate any positive changes.

Conclusions: Studies of the effect of teaching critical appraisal skills on gains in knowledge at the undergraduate level showed consistent improvement. By contrast, changes in knowledge at the residency level were small. Several suggestions from the educational literature are offered to increase effectiveness of critical appraisal interventions.

Résumé

Objectif : Examiner les données probantes selon lesquelles l'enseignement de techniques d'évaluation critique (médecine fondée sur des données probantes) aux étudiants en médecine de premier cycle ou aux résidents entraînera des gains importants de connaissances et l'utilisation accrue de la littérature scientifique dans la prise de décisions cliniques.

Sources de données : Articles publiés de 1966 à 1995 extraits dans le cadre d'une recherche effectuée dans MEDLINE et appuyée par des recherches manuelles; examen de bibliographies tenues par des personnes qui enseignent les techniques d'évaluation critique; examen antérieur de méthodologies.

Sélection des études : On a choisi les articles si l'étude portait sur une forme quelconque de groupe témoin, même si la randomisation rigoureuse n'était pas obligatoire, et si l'intervention était suivie d'une mesure du rendement. On a exclu les articles qui décrivaient simplement le processus d'enseignement de techniques d'évaluation critique et utilisaient une forme quelconque «d'indice de bonheur».

Synthèse des données : Il y avait dix études sur l'impact de l'enseignement des techniques d'évaluation critique : six portaient sur des étudiants en médecine et quatre sur des résidents. Les résultats de trois des études étaient presque impossibles à interpréter et on les a donc exclus; ceux des sept autres étaient acceptables sur le plan méthodologique. L'analyse a démontré que des interventions mises en



Evidence

Études

From the Department of
Clinical Epidemiology and
Biostatistics, McMaster
University, Hamilton, Ont.

This article has been peer reviewed.

Can Med Assoc J 1998;158:177-81

‡ See related article page 203



oeuvre dans des programmes de premier cycle ont produit d'importants gains de connaissances, déterminés par un examen écrit (gain moyen, 17,0 %; écart type [ET], 4,0 %). Par ailleurs, les études réalisées au niveau de la résidence ont démontré régulièrement un faible changement du niveau des connaissances (gain moyen, 1,3 %; ET, 1,7 %). Deux études au cours desquelles on a examiné l'utilisation des écrits par des résidents n'ont pu démontrer de changements positifs.

Conclusions : Les études de l'effet que l'enseignement de techniques d'évaluation critique a sur les gains de connaissances au niveau du premier cycle ont démontré une amélioration uniforme. Par ailleurs, les changements ont été faibles au niveau de la résidence. On présente plusieurs suggestions tirées de littérature scientifique en éducation afin d'accroître l'efficacité des interventions d'évaluation critique.

Evidence-based medicine — “the conscientious, explicit and judicious use of current evidence in making decisions about the care of individual patients”¹ — has been recently characterized as “profound enough that it can appropriately be called a paradigm shift.”² Central to the practice of evidence-based medicine is the learning of skills needed to appraise articles critically.³⁻⁵

In a previous review Audet and associates⁶ examined the adequacy of studies that investigated the effectiveness of teaching critical appraisal skills from a methodological perspective using a 17-item checklist. Their primary concern was methodological rigour, and they only briefly examined whether the studies actually demonstrated a positive effect of the intervention. They found that 3 studies showed a significant improvement in knowledge of epidemiology and statistics, that 4 demonstrated a significant change in reading habits and that, in the domain of capacity to read critically, 2 showed a clear enhancement and 2, less convincingly, suggested that students felt more competent.

The observation that many of the studies of critical appraisal contain methodological flaws is a regretful example of the dissociation between “practising” and “preaching.” But this does not, of itself, preclude the demonstration that critical appraisal can or cannot be learned, unless these flaws are sufficiently serious to negate the findings of the studies. The distinction is important. The gains in knowledge in 2 of the 3 studies that Audet and associates used to conclude a significant improvement in knowledge actually showed a small increase: only 3% and 8% respectively; the authors of the third study (a comparison of lecture with slide-tape show) never calculated the change before and after the intervention but claimed that it was significant. Thus, the question of the effectiveness of teaching critical appraisal skills has yet to be subjected to critical scrutiny.

In this article we address the question What is the evidence that instruction is effective in helping students to acquire the knowledge and skills to appraise studies critically and to identify methodological problems?

Methods

We searched the literature for articles describing interventions to teach evidence-based medicine or critical appraisal skills. The inclusion criteria were deliberately broad:

- There had to be some form of control group, although we did not demand a true (i.e., randomized) experimental design. This criterion excluded single-group, before-after study designs.
- There had to be some direct measure of performance in terms of knowledge, skill or self-reported use of the literature. This criterion eliminated a few studies that used student ratings or testimonials as outcome measures.

We used a number of search strategies: MEDLINE was searched for articles published between 1966 and 1995 using the key words “evidence-based medicine,” “critical appraisal” and synonyms (e.g., “critical thinking”), and “education” or “teaching.” This search was supplemented by a manual search of the reference lists of the articles retrieved, theses and bibliographies maintained by individuals involved in teaching critical appraisal skills, as well as discussions with experts.

In total, before application of the inclusion and exclusion criteria, we located 17 original studies. However, 4 were simple descriptions of a course in critical appraisal skills or epidemiology accompanied by an evaluation based on a “happiness index” and were eliminated from further consideration. Another study was eliminated because it used a before-after design. This left 12 articles that used some form of experimental design and an objective evaluation of performance at the end of the instruction;⁷⁻¹⁸ however, 2 of these studies^{7,8} were comparisons of 2 instructional methods and were eliminated.

A description of the 10 studies is shown in Table 1. Six involved medical students, and 4 involved residents. All used 2-group (treatment and control) designs. One study used historical controls, one was a cohort study, and another used a crossover design. All of the studies used some form of written test to assess either knowledge or skills, or both, made up of 4 to 20 multiple-choice, true/false or short-answer questions. Three of the studies assessed self-reported use of the literature.



All of the undergraduate studies took place in the internal medicine clerkship and appeared to be part of course credit. The interventions varied in length from 3 to 16 hours. In one study¹² the subjects were tutors, who then presumably introduced the concepts from the intervention in their tutorials.

All of the studies involving residents used a variation of the journal-club format: residents were asked to read and appraise articles critically under the guidance of an expert. The sessions typically lasted about an hour once per week. The mean number of sessions attended by each resident ranged from 5 to 17.

Results

We applied the criteria used by Audet and associates⁶ to the studies in our analysis: 3 of the 10 studies had methodological scores below 35%; the remaining 10 had scores

above 50% (highest 83%). Results from the 3 “failing” studies were nearly uninterpretable and were omitted from further analysis. The 7 “passing” studies appeared to have relatively minor methodological problems. All used random allocation or an allocation strategy that was unrelated to the intervention and hence was unlikely to be biased, had complete follow-up of subjects and used appropriate statistical analysis.

For these 7 studies we calculated the mean change in score across all the outcome measures and then the mean difference and standard deviation (SD) over all studies. We found considerable consistency in the findings of these studies.

Medical students

We found consistent evidence that critical appraisal

Table 1: Summary of studies assessing effectiveness of teaching critical appraisal skills to medical students and residents

Study	Design*	Sample†	Intervention	Outcome measure‡	Outcome§	Overall difference, %
Students						
Heller et al ⁹	CT	E (-) U6 C (-)	1.5 h/wk × 12 wk	MCQ (15)	E: 26% “lower” C: 52% “lower”	NA
Radack et al ¹⁰	CT	E (22) U4 C (22)	50 min/wk × 5 wk	Article critique	E: 77% improvement C: 75% improvement	NA
Riegelman ¹¹	CT	E (91) U1 C (82)	16 h; lecture, seminar	MCQ (4)	E: 90% ¶ C: 76%	14
Bennett et al ¹²	CT	E (45) U4 C (34)	16 h; tutorial	Article critique	E: Dx 58%, Mx 35% C: Dx 27%, Mx 23%	21
Frasca et al ¹³	CT	E (48) U3 C (44)	15 h; lecture, seminar	MCQ (20)	E: 49% ¶ C: 28%	21
Landry et al ¹⁴	RCT	E (65) U4 C (65)	3-h seminar	MCQ (10) before and after; problem write-up	E: 82% ¶ C: 74%	12
Residents						
Gehlbach et al ¹⁵	Historical controls; cohort	E (23) R2,3 C (12) R1	8 h	MCQ (7)	E: 74% C: 64%	NA
Linzer et al ¹⁶	RCT	E (42) R2,3 C (43)	Journal club 1/wk × 1 yr	MCQ (15); no. of articles/mo	E: 61% C: 62%	-1
Linzer et al ¹⁷	RCT	E (22) R1 C (19)	Journal club, conference 1/wk × 1 yr	MCQ (15) before and after; no. of articles/mo	Kn: E +10%, C +2% ¶ Sk: E +3%, C +10%	2
Kitchens et al ¹⁸	CT; crossover	E (51) R1-3 C (32)	Reading, seminar; 1st arm 17 wk 2nd arm 12 wk	MCQ (22)	E/C: 66% (-1.4%) C/E: 63% (+5.2%) ¶	3

*CT = controlled trial (near-random assignment), RCT = randomized controlled trial.

†E = experimental group (sample size in brackets), C = control group (sample size in brackets), U = undergraduate (with year), R = resident (with year).

‡MCQ = multiple-choice questionnaire (number of questions in brackets).

§Dx = diagnosis, Mx = management, Kn = knowledge, Sk = skills.

||Overall difference in percentage between treatment and control groups averaged across outcome measures. NA = not applicable.

¶ Statistically significant difference ($p < 0.05$).

courses of several weeks' duration, involving a variety of formats, substantially increased medical students' level of knowledge of topics in clinical epidemiology, as assessed by some form of written test. The mean difference in score across all studies was 17.0% (SD 4.0%); this difference was both educationally and statistically significant.

One of the studies¹⁴ that showed a significant increase in knowledge following the course explicitly asked students in both the treatment and control groups to apply their knowledge of the literature to patient write-ups. At the end of the course, students in both groups cited literature in 53% of their write-ups, but less than 3% in either group mentioned methodological quality. Thus, although students may gain knowledge of methodological topics, this is no guarantee that the knowledge will be used to examine the literature critically.

Residents

In contrast to the medical students, the residents had small gains in knowledge, although reported as statistically significant in 2 of the 3 studies. Overall, the mean difference was 1.3% (SD 1.7%). In one study¹⁷ although the experimental group showed a small gain in knowledge relative to the control group (10% v. 2%), the control group demonstrated a comparable advantage in learning critical appraisal skills (10% v. 3%). Another study¹⁸ was seriously compromised by the use of a crossover design. After the first period, there was no difference between the groups; after the crossover, there was a significant difference between the groups, but this emerged from the combination of a loss of 1.4% in the experimental/control group and a gain of 5.2% in the control/experimental group. The finding of significant differences in the second period is irrelevant, since educational interventions cannot use a washout period.

Consistent with the findings from the studies involving medical students, the 2 studies that examined reading habits of residents found no evidence of an increased use of the literature or a more critical approach to journal articles. In one,¹⁷ residents in both the experimental and the control groups read about 2 articles per month fewer following the intervention. In the other,¹⁶ residents in the intervention group read significantly fewer articles than those in the control group (14.9 v. 23.5 per month). Although in both studies the residents in the experimental groups said that they were more critical in their reading, this was not reflected in the objective tests of knowledge, the results of which were the same¹⁶ or only slightly better¹⁷ following the intervention.

Discussion

Our review clearly demonstrates that, although instruc-

tion in critical appraisal (evidence-based) skills can result in sizeable gains in knowledge among students, the effect of such instruction is much smaller among residents. Furthermore, the minimal evidence to date does not, as yet, provide any indication that the gains in knowledge result in a change in behaviour with respect to the critical use of the literature.

What is the explanation for these findings? The lack of difference among residents cannot be explained by design problems such as co-intervention or contamination, since these would lead to large, not small, gains in knowledge in both the experimental and control groups. There may have been problems with the outcome measures, particularly with the use of brief questionnaires, but this is more an issue of external validity or relevance than of bias. The small sample in many of the studies was a limitation and may have led to low power to detect differences. However, it is not simply that the results were, for the most part, nonsignificant; rather, the results for the residents were consistently close to zero, even when they were statistically significant. As Linzer¹⁹ put it:

Although our results were statistically significant, the magnitude of the educational improvement was small. Our experience reveals that, even with a carefully designed curriculum, the fundamentals of clinical epidemiology are not easily taught during residency.

In any case, none of these methodological problems explains the differences we observed between the undergraduate and residency studies. The duration of the interventions were similar in both groups, as were the outcome measures. If we accept that the observed differences are not spurious, what lessons can be learned from the successful undergraduate studies?

One clear difference between the 2 educational levels is the relation between the critical appraisal course and the student evaluation system. In clerkships, it is possible to ensure that part of the course credit is based on performance in the critical appraisal course. It is difficult to determine whether this criterion was met in all the studies reviewed, but in at least 2 of them^{11,13} the course was required for credit; in another¹² the tutor who taught evidence-based medicine was also responsible for final student evaluation. By contrast, the interventions in the residency studies were generally of the journal-club format, with a meeting once a week, but attendance was typically sporadic. No mention was made of how or whether performance in the course was integrated with resident evaluation.

These differences between education levels are critically important, because evaluation has been frequently identified as a major determinant of learning.²⁰ It may be that the integration of evidence-based medicine as an essential and continuing component of a residency program² will show larger and sustained effects that affect patient care. Clearly this would be preferable to the "add-on" nature of the re-



ported interventions to date. However, there is no evidence as yet on the effectiveness of such sustained interventions.

Our findings differ from those of recent reviews of evidence-based medicine,²⁻⁵ whose positive conclusions were based on 2 studies included in our review.^{12,18} The overall findings of Bennett and colleagues¹² are consistent with those from the other undergraduate studies in that the intervention had a relatively large effect on knowledge. However, the study by Kitchens and Pfeiffer¹⁸ involving residents, which used a crossover design, showed no significant difference between the experimental and control groups in the first unbiased comparison, and only a small (although significant) difference in the second. Thus, although we concur with the claim that instruction in critical appraisal can be effective among medical students, we dispute previous claims that evidence exists of its effectiveness among residents.

Finally, the promise of evidence-based medicine remains the potential benefit to patient care. Previous reviews²⁻⁵ included a study involving McMaster University and University of Toronto graduates 10 to 15 years into practice,²¹ which showed a difference of 6% on a multiple-choice test of hypertension knowledge as evidence that instruction in evidence-based medicine can help graduates stay up to date, with a benefit to patient care. However, the 2 cohorts were considerably different at study inception: students at McMaster were selected primarily on personal qualities and those at Toronto were selected mainly according to academic criteria. Furthermore, the 2 interventions (the curricula) differed in many ways other than the instruction in evidence-based medicine: McMaster used a problem-based and small-group approach with emphasis on self-directed learning, and Toronto used a conventional, lecture-based curriculum. Finally, analysis of the test scores that were related to time since graduation showed that the slope of the line in the Toronto group, although negative, was not significantly different from zero, nor was it significantly different from the slope in the McMaster group. The only significant predictor of performance was the school; therefore, a reasonable interpretation is that the observed differences were present at graduation.

Conclusions

Our review indicates that teaching critical appraisal skills can result in significant gains in knowledge of epidemiology in undergraduate programs; however, this knowledge is apparently not applied in clinical practice. Conversely, the evidence to date indicates that teaching such skills in residency programs is ineffective. More intensive programs in instruction² may be more effective in teaching critical appraisal skills, but there is no evidence as yet to substantiate this claim.

On a more fundamental level, although the goal of evidence-based medicine (and, by extension, of teaching critical appraisal skills) is ultimately to improve patient-care decisions by providing practising physicians with tools to keep up to date with current literature, there is as yet no evidence that the gains in knowledge demonstrated in undergraduate critical appraisal courses can be sustained into residency and practice and eventually translated into improved patient outcomes.

This study was supported by a grant from the Educating Future Physicians of Ontario Project, which is funded by Associated Medical Services and the Ontario Ministry of Health.

References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
2. Evidence Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
3. Sackett DL, Rosenberg WMC. On the need for evidence based medicine. *J R Soc Med* 1995;88:620-4.
4. Sackett DL, Rosenberg WMC. On the need for evidence based medicine. *Health Econ* 1995;4:249-54.
5. Sackett DL, Rosenberg WMC. On the need for evidence based medicine. *J Public Health Med* 1995;17:330-4.
6. Audet N, Gagnon R, Ladouceur R, Marcil M. L'enseignement de l'analyse critique des publications scientifiques médicales est-il efficace? Révision des études et de leur qualité méthodologique. *Can Med Assoc J* 1993;148:945-52.
7. Cuddy PG, Elenbaas JK, Coit KJ. The effectiveness of a slide tape program on literature evaluation. *J Biocommun* 1984;11:2-4.
8. Romm FJ, Dignan M, Hermann JM. Teaching clinical epidemiology: a controlled trial of two methods. *Am J Prev Med* 1989;1:50-1.
9. Heller RF, Peach H. Evaluation of a new course to teach the principles and clinical applications of epidemiology to medical students. *Int J Epidemiol* 1984;13:533-7.
10. Radack KL, Valanis B. Teaching critical appraisal and application of medical literature to clinical problem solving. *J Med Educ* 1986;61:329-31.
11. Riegelman RK. Effects of teaching first-year medical students skills to read medical literature. *J Med Educ* 1986;61:454-60.
12. Bennett KJ, Sackett DL, Haynes RB, Neufeld VR, Tugwell P, Roberts RS. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA* 1987;257:2451-54.
13. Frasca MA, Dorsch JL, Aldag JC, Christiansen RG. A multidisciplinary approach to information management and critical appraisal instruction: a controlled study. *Bull Med Libr Assoc* 1992;80:23-7.
14. Landry FJ, Pangaro L, Kroenke K, Lucey C, Herbers J. A controlled trial of a seminar to improve medical student attitudes toward, knowledge about, and use of the medical literature. *J Gen Intern Med* 1994;9:436-9.
15. Gehlbach SH, Bobula JA, Dickinson JC. Teaching residents to read the literature. *J Med Educ* 1980;55:362-5.
16. Linzer M, DeLong ER, Hupart KH. A comparison of two formats for teaching critical reading skills in a medical journal club. *J Med Educ* 1987;62:690-2.
17. Linzer ML, Brown JT, Frazier, LM, DeLong ER, Siegel WC. Impact of a medical journal club on house staff reading habits, knowledge and critical appraisal: a randomized controlled trial. *JAMA* 1988;260:2537-41.
18. Kitchens J, Pfeiffer MP. Teaching residents to read the medical literature: a controlled trial of a curriculum in critical appraisal/clinical epidemiology. *J Gen Intern Med* 1989;4:385-7.
19. Linzer M. Critical appraisal: more work to be done. *J Gen Intern Med* 1990; 5:457-9.
20. Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educ Res* 1995;24:5-11.
21. Shin JH, Haynes RB, Johnston ME. Effect of problem-based, self-directed undergraduate education on life-long learning. *Can Med Assoc J* 1993; 148:969-76.

Reprint requests to: Dr. Geoffrey R. Norman, Rm. 2C14, Health Sciences Centre, McMaster University, 1200 Main St. W, Hamilton ON L8N 3Z5; fax 905 577-0017; norman@fhs.csu.mcmaster.ca