Research article

Critical appraisal skills of family physicians in Ontario, Canada

Marshall Godwin* and Rachelle Seguin

Address: Centre for Studies in Primary Care, Queen's University, Kingston(K7L 5E9), Canada

Email: Marshall Godwin* - godwinm@post.queensu.ca; Rachelle Seguin - seguinr@post.queensu.ca

* Corresponding author

Abstract

Background: Our primary objective in this study was to measure family physicians' knowledge of the key elements that go into assessing the validity and interpreting the results in three different types of studies: i) a randomized controlled trial (RCT); ii) a study evaluating a diagnostic test; and iii) a systematic review (SR). Our secondary objectives were to determine the relationship between the above skills and age, gender, and type of practice.

Methods: We obtained a random sample of 1000 family physicians in Ontario from the College of Family Physicians of Canada database. These physicians were sent a questionnaire in the mail with follow-up mailings to non-responders at 3 and 8 weeks. The questionnaire was designed to measure knowledge and understanding of the basic concepts of critical appraisal skills. Based on the responses to the questions an Evidence Based Medicine (EBM) Knowledge Score was determined for each physician.

Results: A response rate of 30.2% was achieved. The respondents were younger and more likely to be recent graduates than the population of Ontario Family Physicians as a whole. This was an expected outcome. Just over 50% of respondents were able to answer questions concerning the critical appraisal of methods and the interpretation of results of research articles satisfactorily. The average score on the 12-point EBM Knowledge Scale was 6.4. The younger physicians scored higher than the older physicians, and academic physicians scored higher than community-based physicians. Scores of male and female physicians did not differ.

Conclusions: We have shown that in a population of physicians which is younger than the general population of physicians, about 50% have reasonable knowledge regarding the critical appraisal of the methods and the interpretation of results of a research article. In general, younger physicians were more knowledgeable than were older physicians. EBM principles were felt to be important to the practice of medicine by 95% of respondents.

Background

The concept we now refer to as evidence-based medicine (EBM) has its roots in the clinical epidemiology group at McMaster University. In the early and mid 1980s Haynes, Guyatt, Sackett, Oxman and others began writing about how to keep up-to-date by effectively reading and using the medical research literature [1-4]. In the 1990's this same group began publishing "User's guides" in the Journal of the American Medical Association about reading and critically appraising different types of published articles, including articles on therapy [5,6], articles about diagnostic tests [7,8], and systematic reviews [9]. There are

now about 25 user's guides on most of the different types of health related publications. There is a website [10] available at the Centre for Health Evidence that contains the full text of all of these guides.

The concepts of clinical epidemiology, critical appraisal, and evidence-based medicine have permeated through academia over the past decade. These concepts are acknowledged and taught in medical schools and residency programs [11-13]. Still, some doubt its practical application; they worry that it is too much like cookbook medicine and that it denies the importance of experience and patient's wishes [14]. However, EBM is defined in terms of evidence, clinical expertise and patient's values. The accepted definition is "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" or perhaps more clearly it is seen as the "integration of best research evidence with clinical expertise and patient's values" [15]. Best practice does not blindly apply research evidence outside the context of the patient or ignore the filter of clinical expertise.

For a practitioner to be able to interpret research evidence, a very specific body of knowledge and skills is needed. This knowledge, and these skills, are now taught in medical schools and residency programs, and courses are available in many countries for practicing physicians. A number of investigators looked at the effectiveness of courses on evidence based medicine [13,16-19]. These studies generally assess the degree to which new knowledge has been learned and retained or the degree to which behavior has been changed. A systematic review that included six studies on medical students and four on residents was conducted by Norman and Shannon in 1998 [17] to assess the effectiveness of instruction in critical appraisal. They found that medical students showed consistent gains in knowledge from these courses but there was little change for residents. A Cochrane review looking at the effectiveness of teaching critical appraisal skills to health professionals [20] found only one study that met their inclusion criteria and it showed that critical appraisal teaching has positive effects on participants' knowledge. In that, study critical appraisal teaching resulted in a 25% improvement in critical appraisal knowledge compared to a 6% improvement in the control group.

Little has been done to assess the degree to which practicing physicians understand and use critical appraisal skills, or the degree to which they subscribe to the concepts of evidence based medicine. In this study, we are not assessing the effect of a program or course of training, we are assessing the degree to which knowledge, attitudes and skills relating to EBM have permeated into the community of family doctors in Ontario.

Our primary objective in this study was to measure family physicians' knowledge of the key elements that go into assessing the validity and interpreting the results in three different types of studies:

i) a randomized controlled trial (RCT); ii) a study evaluating a diagnostic test; and iii) a systematic review (SR). Our secondary objectives were to determine the relationship between the above skills and age, gender, and type of practice.

Method

We obtained a random sample of 1000 family physicians in Ontario from the 6000 member computerized database of the College of Family Physicians of Canada. The database is capable of generating random samples. These 1000 physicians were sent a questionnaire in the mail with follow-up mailings to non-responders at 3 and 8 weeks. The questionnaire was designed to measure knowledge and understanding of the basic concepts of critical appraisal skills as it applies to the <u>three types</u> of articles listed above.

Measuring Knowledge of Critical Appraisal of Research Methods

To measure the respondents' understanding of the concepts related to critical appraisal we asked them to list two things they would look for when considering the quality of each of the three types of articles. The actual questions were :

• Please list two things you would look for when considering the quality of a journal article reporting on the results of *a randomized controlled trial*

• Please list two things you would look for when considering the quality of a journal article reporting on the results of *a study evaluating a diagnostic test*

• Please list two things you would look for when considering the quality of *a systematic review* of a topic

We asked for just two things, instead of all things, they would look for in order to obtain a sense of what they felt was most important, and to prevent a wild guessing spree of possible items. It also simplified the questionnaire, which was important since we suspected response rate would be an issue.

The respondents were considered to have correctly answered the question regarding *randomized controlled trial* if the following concepts were mentioned: sample size/ power; blinding; base line characteristics; measuring appropriate outcomes; description of randomization;

Table 1: Calculating a number-needed-to-treat(NNT) from the results of a RCT.

A study looked at the effect of a cholesterol lowering drug in patients with coronary heart disease. One outcome the investigators looked at was all-cause mortality. 256 (11.5%) of the 2223 people in the placebo group died during the 5-year follow-up and 182 (8.2%) of the 2221 people in the treatment grouped died during the 5 years. The absolute risk reduction (ARR) was therefore 3.3% or 0.033.
What would the number-needed-to-treat (NNT) be in this case?
NNT = _____

Table 2: Urine dipstick test for diagnosing a urinary tract infection(UTI)

| | | Urine Culture | | |
		Positive	Negative	
Dipstick Test	Positive	450	400	850
	Negative	50	400	450
		500	800	1300

population similar to my patients; intention to treat analysis.

Mention of the following concepts was considered a correct answer for *a study evaluating a diagnostic test*: reference to a gold standard; sensitivity/ specificity/positive predictive value/negative predictive value; likelihood ratios; cost/benefit; base line characteristics; prevalence/likelihood; false positive/false negative; availability of the test in my practice; applicability of the test to my patients.

Of the responses given for *a systematic review*, mention of the following concepts was considered a correct answer: type of articles reviewed; databases searched; specific criteria for search; completeness of the search; inclusion/exclusion criteria for articles; inclusion of published and unpublished data; homogeneity assessment; meta-analysis was done.

Measuring Knowledge of Interpretation of Research Results

To assess understanding of the results from a randomized controlled trial the scenario in Table 1 was presented and the respondent was asked to calculate the NNT. The correct answer is NNT= 30.3 calculated by 1/ARR = 1/0.033. In case the respondent knew how to calculate the answer but did it roughly or rounded up or down, we accepted any answer between 30 and 31.

To assess understanding of the results from *a study evaluating a diagnostic test*, the data in Table 2 were presented and the respondent was asked to: "Please calculate the sensitivity and specificity of the dipstick test and say whether the dipstick test is best for 'ruling in' or 'ruling

out' the presence of a urine infection". The correct answers were sensitivity = 450/500 = 90%, specificity = 400/800 = 50%. The test is best for "ruling out" infection. It is a SnNout meaning that since the test is highly **S**ensitive, a **N**egative test rules **out** a UTI ... or at least nearly rules it out.

To assess understanding of the results from *a systematic review*, the data in Figure 1 were presented and the respondent was asked two questions:

a) Does treatment with antibiotics significantly decrease pain during the first 24 hours after diagnoses of an otitis media? () Yes () No

b) Does treatment with antibiotics significantly decrease pain during days 2 to 7 after diagnoses of an otitis media? () Yes () No

The EBM Knowledge Score
We developed an overall knowledge score for each respondent. For each of the three research methods questions a respondent could get zero, one, or two correct answers. A point was given for each correct answer. Therefore, the highest possible score on the methods section would be six, which would be achieved by getting two correct answers for each of the three questions.

For each of the six research results questions, the possible scores were zero or one, for a possible total of six. The research methods score and the research results score were added for an overall maximum EBM Knowledge Score out of 12.
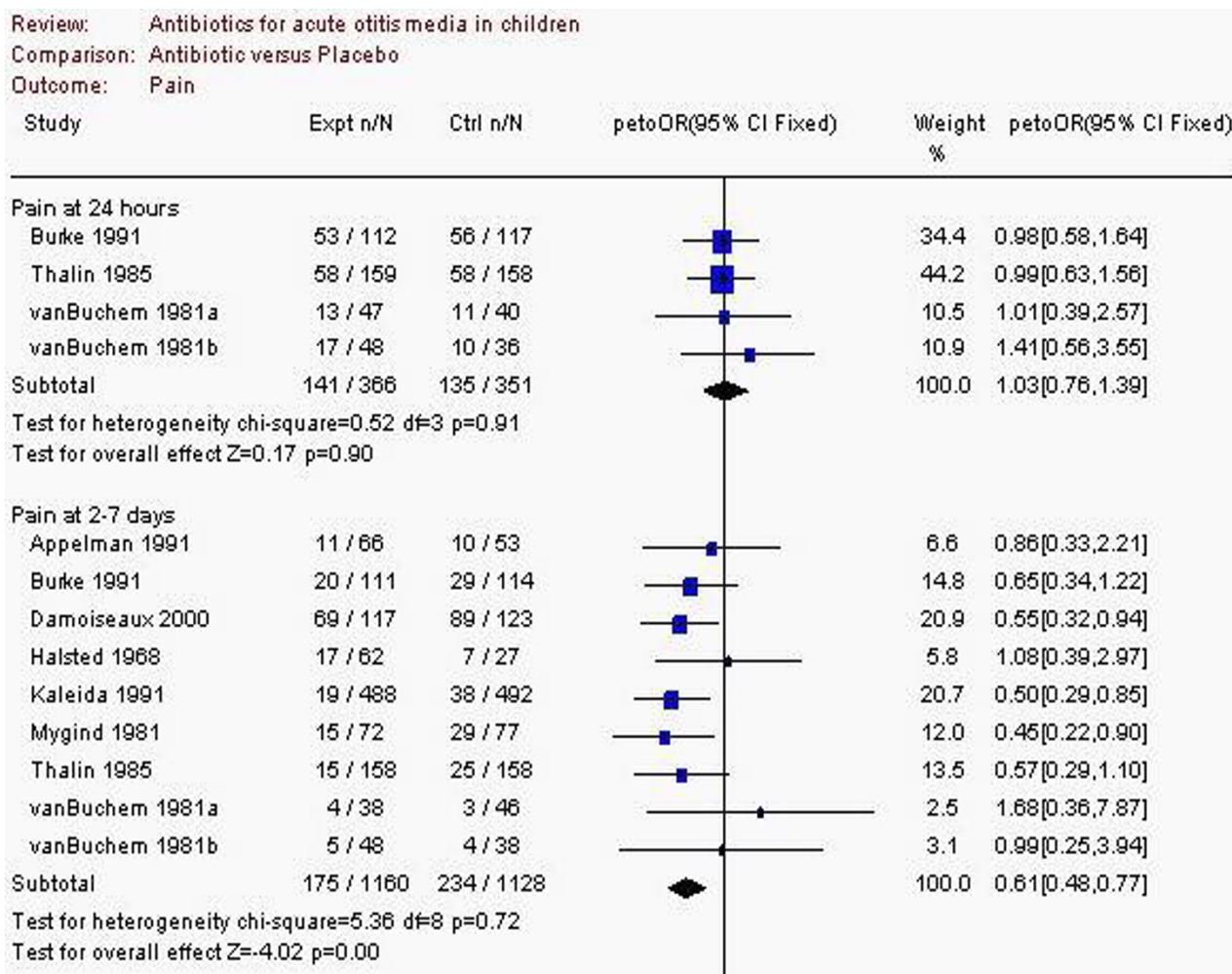
Figure 1
A "forest plot" from a systematic review on antibiotic use in acute otitis media

We also asked demographic questions and questions concerning EBM training and attitudes towards EBM. The study was reviewed and approved by the Queen's University Research Ethics Board.

Results

Of the 1000 family physicians surveyed, 53 were removed from the denominator when calculating response rate for various reasons including wrong address, currently enrolled in a residency program, deceased, and no longer in practice. Of the remaining 947 physicians, 286 (30.2%) responded. The respondents were compared to a sample of Ontario physicians taken from the Janus project data and are presented in Table 3. The Janus project is a survey of Canadian family physicians conducted by the College of Family Physicians every few years. The respondents dif-

fer significantly from the Janus sample and therefore are probably not representative of all Ontario family physicians. We believe this low response rate and the unrepresentativeness of the responders is an important result itself. We will address this in the discussion.

In Table 4, we present the results of the answers to the 12 individual questions. A large majority of respondents knew one correct answer to the methods question for RCT's and studies of diagnostic tests, whereas only 46% could come up with one correct response for critically appraising the methods of a systemic review. Interestingly, respondents were better able to interpret the results of systematic reviews than for RCT's or studies of diagnostic tests. The ability to calculate NNT and sensitivity and specificity was only in the range of 50% (47 – 57%) even

Table 3: Comparison of demographics of survey respondents with a sample of Janus data

	EBM sample N = 286	Janus sample N = 1385	p-value
Sex (Female)	128 (45%)	482 (35%)	.007
Age group			
25–35	76 (27%)	162 (12%)	<.001
36–45	115 (40%)	424 (31%)	
46–55	74 (26%)	463 (34%)	
55–65	21 (7%)	336 (24%)	
Graduation Year	1986 (SD 8.8 years)	1979 (SD 11.2 years)	<.001
Academic	24 (8.4%)	68 (5%)	<.001

Table 4: Percent of Correct Responses to Individual Questions

	At least one correct answer	Two correct answers
RCT methods	242 (85%)	126 (44%)
Diagnostic test methods	210 (70%)	129 (45%)
Systematic review methods	132 (46%)	51 (18%)
	Correct Answer	
NNT	137 (48%)	
Sensitivity	164 (57%)	
Specificity	137 (47%)	
Test best for ruling out?	123 (43%)	
SR: Pain decrease 24 hrs	202 (71%)	
SR: Pain decrease 2–7 days	184 (64%)	

when the actual numbers required to do these calculations were given up front.

In Table 5, we present the mean and standard deviations of EBM scores for various subgroups. The mean score was only slightly above 50% (6.4 out of 12) for the whole group. Men and women performed equally well. However, as age increased the mean score decreased. While family physicians in academic settings scored better than physicians in community settings did, one would have hoped for an even better performance from the academic physicians who teach these concepts to family medicine residents.

Table 6 contains the responses to questions on EBM training and attitudes. About half the respondents had received critical appraisal instruction in medical school or residency; a quarter had sought out continuing medical education courses on EBM skills since graduating; just over half felt they had moderate to high EBM skills; and a large majority (95.5%) believed EBM principles were important to the practice of medicine.

Discussion

Our main objective with this study was to measure family physicians' knowledge and skill level regarding critical appraisal. Our first discovery, as indicated by the relatively low (30.2%) response, was that physicians did not want to respond to our questionnaire. This might have been predicted since physicians do not like responding to questionnaires that measure knowledge, especially if it is about a topic they do not know a lot about. The poor response rate is probably itself a proxy measure of physicians' level of knowledge with the topic. Physicians who felt comfortable with the topic were more likely to respond. Our results then should be considered as the best-case scenario; the truth for the whole population of physicians in Ontario is probably that they are overall less knowledgeable about critical appraisal then our study would suggest. What we see in this study is "as good as it gets."

This interpretation is upheld by the comparison of our respondent population with the Janus data. Our respondents were younger and more likely to be female, which fits with later years of graduation. While our population covered the full age range of physicians, it was overall a younger group then currently makes up the physician

Table 5: EBM Knowledge Scores (out of maximum of 12) for the whole group and by subgroups.

		Mean Score (Standard deviation)	P value
Whole Respondent Group N = 286		6.4 (3.06)	NA
Sex	Female	6.2 (3.2)	NS
	Male	6.5 (3.0)	
Age	25–35	8.2 (2.8)	0.001 chi sq for trend
	36–45	6.2 (2.7)	
	46–55	5.4 (2.9)	
	56–65	4.4 (3.0)	
Practice Base	Academic	8.0 (2.6)	0.007
	Community	6.0 (3.5)	

Table 6: Training and Attitudes

Instruction in medical school on EBM	151 (52.8%)
Instruction in residency on EBM	164 (57.3%)
Attended CME to learn EBM skills	73 (25.5%)
Moderate or high critical appraisal skills	157 (54.9%)
EBM principles are very* important or moderately* important	273 (95.5 %)

* 140(49%) **133(46.5%)

population in Ontario. For these young physicians, evidence based medicine and critical appraisal were part of the curriculum throughout their training. More of them probably felt comfortable responding to the questionnaire. Despite this, the results are not impressive. While they almost unanimously agree that EBM is important, the overall EBM Knowledge Score was just 6.4 out of 12 for the whole group and even the recent graduates (age 25–35 years old group) achieved only 8 out of 12 on the overall EBM Knowledge Scale. They probably would have considered a mark of 67% a disaster during their training!

The academic family physicians did significantly better than their community-based colleagues, but the highest score was 8.5.

One could put a positive spin on this data. Table 4 suggests that approximately 50% of physicians have a reasonable grasp of critical appraisal concepts and know how to interpret research results; half of them can calculate an NNT and sensitivity and specificity and know what they mean. If we had done this survey 10–15 years ago, it is likely the results would have been much different. The EBM and critical appraisal teaching of the past decade has undoubtedly had an effect.

There are several limitations to this study. The low response rate is the main limitation. However, we are able to know how the low response rate makes the demographics of our study population different from the population as a whole by comparing it to the Janus data. This allows us to consider these differences when interpreting the results. Our data over-estimates the critical appraisal knowledge level of physicians in Ontario. The other limitation with this type of survey is that the physicians might have asked a colleague or sought information from a book or online to answer the questions. While this may have resulted in their learning more about critical appraisal because of this survey, it would skew the results to appear better than they really are. A final limitation is that the EBM Knowledge Score we devised from the questions in the survey has not been tested for reliability or validity. The degree to which a score on this scale is the same if repeated, and the degree to which a score on this scale truly measures one's knowledge of EBM or critical appraisal is not clear.

Conclusion

We have shown that in a population of Ontario physicians that is younger than the general population of physicians, about 50% have reasonable knowledge regarding the critical appraisal of the methods and the interpretation of results of a research article. In general, younger

physicians were more knowledgeable than older physicians were. EBM principles were felt to be important to the practice of medicine by 95% of respondents.

List of Abbreviations

EBM = Evidence Based Medicine

RCT = Randomized Controlled Trial

SR = systematic review

NNT = Number Needed to Treat

ARR = absolute risk reduction

Competing Interests

None declared.

Authors' Contributions

Both authors have reviewed and approved the text of the manuscript. All authors contributed to 1) the conception and design or analysis and interpretation AND 2) the initial drafting or critical revision of the content.

Acknowledgements

References

1. Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL: **How to keep up with the medical literature: I. Why try to keep up and how to get started.** *Ann Intern Med* 1986, **105(1):**149-153.
2. Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL: **How to keep up with the medical literature: II. Deciding which journals to read regularly.** *Ann Intern Med* 1986, **105(2):**309-312.
3. Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL: **How to keep up with the medical literature: III. Expanding the number of journals you read regularly.** *Ann Intern Med* 1986, **105(3):**474-478.
4. Oxman AD, Guyatt GH: **Guidelines for reading literature reviews.** *CMAJ* 1988, **138(8):**697-703.
5. Guyatt GH, Sackett DL, Cook DJ: **Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group.** *JAMA* 1993, **270(21):**2598-2601.
6. Guyatt GH, Sackett DL, Cook DJ: **Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group.** *JAMA* 1994, **271(1):**59-63.
7. Jaeschke R, Guyatt GH, Sackett DL: **Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group.** *JAMA* 1994, **271(9):**703-707.
8. Jaeschke R, Guyatt G, Sackett DL: **Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group.** *JAMA* 1994, **271(5):**389-391.
9. Oxman AD, Cook DJ, Guyatt GH: **Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group.** *JAMA* 1994, **272(17):**1367-1371.
10. **Centre for Health Evidence User's Guides to Evidence Based Practice.** 2003.
11. Fliegel JE, Frohna JG, Mangrulkar RS: **A Computer-based OSCE station to measure competence in evidence-based medicine skills in medical students.** *Academic Medicine* 2002, **77(11):**1157-1158.
12. Smith CA, Ganschow PS, Reilly BM, Evans AT, McNutt RA, Osei A *et al.*: **Teaching evidence-based medicine skills: A controlled trial of effectiveness and assessment of durability.** *J Gen Intern Med* 2000, **15:**710-715.
13. Linzer M, Brown JT, Frazier LM, DeLong ER, Siegel WC: **Impact of a medical journal club on house-staff reading habits, knowledge, and criteria appraisal skills: A randomized control trial.** *JAMA* 1988, **260:**2537-2541.
14. Mayer J, Piterman L: **The attitudes of Australian GPs to evidence-based medicine: a focus group study.** *Fam Pract* 1999, **16(6):**627-32.
15. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes WB: **Evidence-Based Medicine: How to Practice and Teach EBM.** *Churchill Livingston* .
16. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, Kunz R: **Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine.** *BMJ* 2002, **325:**1338-1341.
17. Norman GR, Shannon SI: **Effectiveness of instruction in critical appraisal (evidence based medicine) skills: a critical appraisal.** *CMAJ* 1998, **158:**177-181.
18. Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R: **A systematic review of the effectiveness of critical appraisal skills training for clinicians.** *Med Educ* 2000, **34:**120-125.
19. Ibbotson T, Grimshaw J, Grant A: **Evaluation of a programme of workshops for promoting the teaching of critical skills.** *Med Educ* 1998, **32:**486-491.
20. Parkes J, Hyde C, Deeks J, Milne R: **Teaching critical appraisal skill in health care settings.** *The Cochrane Database of Systematic Review* 2002.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6920/3/10/prepub