Prescribers' Knowledge and Skills for Interpreting Research Results: A Systematic Review

Leila Kahwati, MD, MPH[1], **Dennis Carmody, MPH**[1], **Nancy Berkman, PhD**[1], **Helen W. Sullivan, PhD, MPH**[2], **Kathryn J. Aikin, PhD**[2], **Jessica DeFrank, PhD**[1]

[1]RTI International, Research Triangle Park, NC

[2]Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD

Abstract

Introduction: Appropriate medication prescribing may be influenced by a prescriber's ability to understand and interpret medical research. The objective of this review was to synthesize the research related to prescribers' critical appraisal knowledge and skills—defined as the understanding of statistical methods, biases in studies, and relevance and validity of evidence.

Methods: We searched PubMed and other databases from January 1990 through September 2015. Two reviewers independently screened and selected studies of any design conducted in the United States, the United Kingdom, or Canada that involved prescribers and that objectively measured critical appraisal knowledge, skills, understanding, attitudes, or prescribing behaviors. Data were narratively synthesized.

Results: We screened 1,204 abstracts, 72 full-text articles, and included 29 studies. Study populations included physicians. Physicians' extant knowledge and skills were in the low to middle of the possible score ranges and demonstrated modest increases in response to interventions. Physicians with formal education in epidemiology, biostatistics, and research demonstrated higher levels of knowledge and skills. In hypothetical scenarios presenting equivalent effect sizes, the use of relative effect measures was associated with greater perceptions of medication effectiveness and intent to prescribe, compared with the use of absolute effect measures. The evidence was limited by convenience samples and study designs that limit internal validity.

Discussion: Critical appraisal knowledge and skills are limited among physicians. The effect measure used can influence perceptions of treatment effectiveness and intent to prescribe. How critical appraisal knowledge and skills fit among the myriad of influences on prescribing behavior is not known.

Corresponding author: Leila Kahwati, MD, MPH, RTI International, 3040 E. Cornwallis Road, Research Triangle Park, NC 27709, Tel: (919) 541-6629, Fax: (919) 541-6683, lkahwati@rti.org.

Safe and appropriate prescribing of medications is influenced by numerous factors. Differences in understanding and perception of medication effectiveness by prescribers may lead to different therapeutic decisions, even with accurately presented data. A recent systematic review found that when it comes to prescribing new medications, several factors, such as being male, being younger, and the location of a prescriber's training, all increased the likelihood of prescribing new medications.[1] This review also identified that factors related to scientific orientation play a role in prescribing behavior. The number of peer-reviewed articles read, attendance at continuing medical education courses, and perceived scientific orientation (e.g., physicians who valued staying current with scientific developments more than spending time with patients) were associated with a higher likelihood of prescribing new medications.[1] A clinician's ability to correctly interpret research evidence to guide therapeutic decision making is essential to reducing unwarranted variation and clinically inappropriate prescribing.

Evidence-based medicine (EBM) is the judicious use of the best current evidence in making decisions regarding patient care.[2] During the last two decades, EBM has been introduced into the U.S. health professional training and practice community.[2] A core concept of EBM is critical appraisal of evidence (i.e., research studies), which is defined as the understanding of scientific and statistical methods, the ability to identify biases in studies, and the ability to determine if the evidence is relevant and valid and how it affects patient care.[3] The American Association of Medical Colleges (AAMC) includes the domain *Knowledge for Practice* as one of eight competencies comprising its list of common learner expectations.[4] This competency states, "Demonstrate knowledge of established and evolving biomedical, clinical, epidemiological, and social-behavioral sciences, as well as the application of this knowledge to patient care."

Although EBM has been adopted in recent decades, more than half of actively licensed physicians are currently aged 50 or above[5] and were trained before the current EBM era. Studies have reported that physicians lack sufficient critical appraisal knowledge and skills. [6–9] Because of this, prescribers may be unduly influenced by the way study results are presented.

Although many prescribers may lack sufficient critical appraisal knowledge and skills, many believe it is essential for clinical practice and are aware of the gaps in their knowledge and need for more education or training. One study of 300 teaching faculty, medical residents, and medical students found that 88% believed that EBM is important for clinical practice and 93% affirmed that biostatistics is an important part of EBM.[10] However, only 9% felt that they have had adequate training in biostatistics, and only 23% believed that they could identify whether the correct statistical methods have been applied in a study.[10] Another study of 317 physicians in the United Kingdom (U.K.) found similar results.[11] In this study, physicians rated that they were highly confident that "EBM is essential" (mean = 5.10 on 6-point Likert scale) but were also confident that "I need more training" (mean = 5.30).

The objective of this systematic review was to summarize the research related to prescriber critical appraisal knowledge and skills. Specifically, we defined two key questions:

Key Question 1: Do prescribers' education, training, and skills related to the critical appraisal of research influence their ability to correctly interpret research and influence their behavior in prescribing medications?

Key Question 2: What practices of research presentation are associated with correct interpretation of research by prescribers?

Methods

With the assistance of a medical librarian, we searched PubMed, Cochrane Library, PsycINFO, and ClinicalTrials.gov for original research published in English, from January 1990 through September 2015, using the search strategy in Appendix A1 in the Digital Supplement. We identified additional studies by hand-searching reference lists of relevant studies and forward-tracing relevant studies using Google Scholar and Web of Science.

Study Selection

Two reviewers independently screened titles/abstracts and full-text articles for inclusion based on study selection criteria detailed in Appendix A2. Disagreements at the full-text review stage were resolved by discussion. In brief, we included all study designs involving prescribers, which we defined as physicians, nurse practitioners, and physician assistants. We considered only studies that took place in the United States (U.S.), the U.K., and in Canada because these countries are where most evidence-based medicine concepts and training have originated. For Key Question 1, we included studies of experimental or nonexperimental exposures involving critical appraisal training or education, which we defined as courses, trainings, or other formal experiences with clinical epidemiology, biostatistics, and evidence-based medicine, with objectives to measure or influence prescribers' ability to read, interpret, and apply medical research studies. For Key Question 2, we included studies evaluating alternative result formats, including framing of subgroup and post hoc analyses. Eligible outcomes included the correct interpretation of research results through objective critical appraisal knowledge or skills assessment and prescribing attitudes and behaviors.

Data Abstraction and Synthesis

We developed a standardized abstraction form to capture relevant study information. One reviewer abstracted data from included studies, and these data were checked for accuracy by a senior reviewer. We contacted study authors when needed for further clarification. The senior reviewer evaluated study quality by assessing the potential for selection bias, performance bias, and measurement bias and whether appropriate analytic techniques were employed; this assessment was tailored to the specific study design. We narratively synthesized findings for each key question by summarizing the characteristics and results of included studies in narrative format. We did not quantitatively synthesize findings because of methodological and clinical heterogeneity.

Results

We screened 1,204 titles and abstracts and 72 full-text articles that we identified through our electronic database search and supplemental hand search. We included 29 unique studies for this review (Figure 1). We describe individual study characteristics, outcomes, study quality, and applicability of included studies in the evidence tables located in Appendix B in the Digital Supplement.

Key Question 1

Do prescribers' education, training, and skills related to the critical appraisal of research influence their ability to correctly interpret research and influence their behavior in prescribing medications?

Study Characteristics—Of the 19 studies relevant to this key question, two were systematic reviews,[12,13] three were randomized controlled trials (RCTs),[14–16] two were nonrandomized trials,[17,18] five were repeated measures designs using a single group with one pre- and one or more post-intervention measurements,[19–23] six were cross-sectional designs,[24–29] and one was a psychometric validation study.[30] Except for one study that was published in 1980,[29] all studies were published in the year 2000 or later. All studies were focused on physician prescribers. Figure 2 summarizes the biostatistical, epidemiological, and critical appraisal concepts evaluated among these studies. Intervention studies evaluated stand-alone teaching interventions, such as journal clubs, seminars, and workshops in addition to teaching approaches integrated into clinical practice. Interventions ranged in duration from several hours to longitudinal curricula over 2 years.

Most studies reported outcomes related to objectively measured critical appraisal knowledge or skills. Depending on study design, study authors reported these outcomes as a change from pre-intervention to post-intervention within a single group (repeated measures designs), a mean change in knowledge in an intervention group as compared with a control group (RCTs, nonrandomized trials), or as differences in knowledge among groups characterized by differences in demographics, prior education and training, or other study population characteristics (cross-sectional designs). Four studies used existing EBM assessments (Fresno test[20–22]; Berlin EBM questionnaire[17]). All other studies used assessments developed for the study; eight of these provided information about the assessment validity or reliability.[14–16,18,25,27,28,30]

Findings—Table 1 provides a summary of findings for the 16 studies for which extant critical appraisal knowledge or skills were reported.

Knowledge—The mean percent correct on critical appraisal knowledge assessments (which varied in length and rigor) ranged from 34% to 74% among the five studies using cross-sectional designs.[24,25,27–29] Several characteristics related to scientific orientation were found to be associated with higher knowledge. These include having an advanced degree (other than a medical degree) or prior training in epidemiology, biostatistics, or clinical research.[27–29] Residents and faculty in university settings or academic positions

were more likely to have significantly higher knowledge scores than residents and faculty in community settings.[24,25,28]

One systematic review and five primary research studies reported outcomes related to objectively measured changes in knowledge in response to an intervention/exposure.[12,14,16–19] Coomarasamy and Khan[12] synthesized findings from 17 studies reporting knowledge; 12 of the included studies demonstrated improvements in knowledge. Of the four primary research studies with control groups, two showed statistically significant improvements in knowledge scores compared with the control group,[16,17] one showed no effect of the intervention,[14] and one showed favorable effects for one of the two active study arms evaluated.[18]

Skills—Two studies reported on extant critical appraisal skills.[25,26] Beasley and Wooley[25] reported mean critical appraisal assessment scores of between 0.6 and 1.9 on a 4-point scale among statewide faculty affiliated with a single U.S. medical school. Mimiaga et al.[26] assessed the ability of 115 specialists and generalist physicians to correctly interpret efficacy findings from two trials related to pre-exposure prophylaxis. In this study, 72% of respondents correctly interpreted trial results.

Two systematic reviews and seven primary research studies reported outcomes related to objectively measured changes in critical appraisal skills.[12,13,15,16,20–23,30] Within the Coomarasamy and Khan[12] review, five of the nine studies assessing skills reported an increase and four reported no increase in skills. Within the Harris et al.[13] review, five of the seven studies that assessed skills reported statistically significant increases.[13] Of the studies that demonstrated increases, three included a mentoring component, four included didactic support, and four used a structured review instrument. Of the three primary research studies with control groups, one showed no differences in skills in two of the three domains assessed,[16] and the other two studies showed improvements in skills in the intervention groups.[15,30] All four primary research studies using repeated measures designs showed statistically significant improvements in skills post-intervention.[20–23]

Attitudes—Five studies, including one systematic review, reported on prescriber attitudes, which included self-assessed confidence in understanding specific concepts and in interpreting evidence, as well as self-acknowledgement of the need for certain knowledge or skills for different clinical scenarios.[12,16,21,27,28] Within the Coomarasamy and Khan[12] review, attitude outcomes were assessed by six intervention studies of which half showed statistically significant improvements relative to a comparison group. Among plastic surgery residents from a single U.S. program, Susarla et al. found that residents with a prior course in biostatistics or EBM reported more confidence in their ability to interpret the results of statistical tests than residents without formal coursework; performance on knowledge assessments was strongly correlated with resident confidence.[27] Windish et al. reported a mean confidence score of 11.7 (SD 2.7) out of 20 for interpreting statistical results in survey of internal medicine residents from 11 programs in one U.S. state.[28] In this study, 75% of respondents reported not understanding all of the statistics encountered in the literature. Taylor et al.[16] reported confidence outcomes in an RCT evaluating a 3-hour EBM workshop conducted among practitioners in the U.K., using six items rated on a 5-point Likert item

scale for a possible score range of 6 to 30. Post-intervention mean scores were 15.0 (SD 5.3) in the intervention group and 13.8 (SD 5.1) in the control group, with no statistical difference between groups. Conducted in a single U.S. family medicine residency program, Shaughnessy et al. used a repeated measures design to report participant confidence in their ability to determine five characteristics of a study. The mean confidence score pre-intervention was 17.9 (95% confidence interval [CI] 16.6 to 19.3), out of a maximum score of 25, and the mean score post-intervention was 21.1 (95% CI 19.5 to 22.7).

Prescribing Behavior Intentions—One study reported outcomes related to prescribing behavior intentions. Using a cross-sectional design, Mimiaga et al.[26] assessed physicians' characteristics, critical appraisal skills, and likelihood of prescribing based on data presented in two clinical trials focused on pre-exposure prophylaxis for HIV infection. Between 78% and 83% indicated they would prescribe treatment to populations in which efficacy had been demonstrated in the trials; just over half (52 to 53%) also indicated they would prescribe to populations that were not included in the trials.

Key Question 2

What practices for research presentation are associated with correct interpretation of research by prescribers?

Study Characteristics—Of the ten studies relevant to this key question, two were systematic reviews,[31,32] one was an RCT,[33] one was a factorial RCT,[34] five were repeated measures design with a single post-intervention measurement,[35–39] and one was a qualitative research design.[40] Three studies were published between 1990 and 2000,[33,35,36] and seven published after 2000.

Both systematic reviews reported on studies conducted with physicians, consumers, and non-prescribing health care professionals. We generally limited our synthesis to the subset of studies within each review that focused on physicians, except where an outcome was not presented separately for physicians. Of the primary research studies, six were conducted solely among physicians in practice (i.e., beyond residency training),[3,34,35,38–40] and two included a mixed study population of residents and physicians in practice.[33,36]

Studies used actual or simulated data presented to participants to assess the impact of alternative effect-measure presentations. For example, a common intervention across numerous studies was to present the same efficacy data for a treatment using relative risk reduction (RRR) versus using absolute risk reduction (ARR). All primary studies were designed as experiments involving a one-time exposure to the actual or simulated data, followed immediately by an assessment. Outcomes varied but included measures designed to assess participant understanding of the data presented,[32,34] rating of treatment effectiveness,[32,33,31] and likelihood of making a decision to prescribe.[31,32,35–39] In the single included qualitative research study, the outcomes assessed were usefulness and clarity of reporting methods, preference for reporting, and attitudes toward reporting methods.[40]

Findings

Perceived Effectiveness of Treatment: The Akl et al.[32] review included three studies assessing perceived effectiveness of treatment by health care professionals. Despite numerically equivalent results, professionals rated interventions as being more effective when the result was reported as RRR compared with when the result was reported using number needed to treat (NNT) (pooled standardized mean difference [SMD] 1.15, 95% CI 0.80 to 1.50). Similarly, professionals perceived higher effectiveness when studies reported results with RRR compared with ARR (pooled SMD 0.39, 95% CI −0.04 to 0.82), but these findings were not statistically significant. Professionals also perceived higher effectiveness when studies were reported using ARR compared with studies reported using NNT (pooled SMD 0.79, 95% CI 0.43 to 1.15). The Covey et al.[31] review found that across all audience types, numerically equivalent results presented in a relative format had statistically significant higher ratings of effectiveness than results presented using an absolute format (log odds 1.60, 95% CI 1.60 1.22 to 1.98, N = 29 comparisons) or results presented using NNT (log odds 1.64, 95% CI 1.28 to 2.00, N = 24 comparisons).

Using a factorial RCT design Raina et al.[34] randomized 120 Canadian physicians identified from membership rolls of a medical professional society to receive one of six case scenarios involving data from a meta-analysis to determine how effect-measure used, disease severity, magnitude of effect, and statistical consistency of included studies affect physician interpretation of treatment effect. The authors reported no difference in perceived treatment effectiveness by presentation of relative (i.e., risk ratio, odds ratio) versus absolute measures (i.e., risk difference) or by severity of the disease for which treatment was being evaluated. Physicians rated data as having higher levels of treatment effectiveness when the effect size was large, rather than when the effect size was small. Further, physicians rated data as having higher levels of treatment effectiveness when findings were statistically consistent than when they were statistically inconsistent; and statistical consistency affected this perception more when the effect-measure used was risk difference.

Intent to Prescribe: Six studies evaluated the impact of relative versus absolute data presentation formats on intent to prescribe.[32,35–39] The Akl et al.[17] review found that among 12 studies, interventions using RRR were rated as being more persuasive for making a decision to treat or adopt the intervention (pooled SMD 0.71, 95% CI 0.49 to 0.93) compared with intervention using ARR, despite numerically equivalent results. Similarly, this review found that among 10 studies, interventions using RRR were also rated as more persuasive (pooled SMD 0.65, 95% CI 0.42 to 0.87) compared with studies using NNT, but no difference in persuasiveness was seen among studies reporting ARR versus NNT. Four primary research studies, all using repeated measures designs, evaluated the intent to prescribe based on whether relative or absolute measures were used.[35–38] Across these studies, numerically equivalent results presented using a relative format resulted in significantly higher willingness to prescribe compared with an absolute format.

Using a repeated measures design, Parker et al.[39] evaluated patient management decisions for six simulated clinical scenarios among 435 physicians in a single Canadian province. The scenarios were designed to assess differences in management decisions based on

whether overall or subgroup benefits or harms were present and whether a treatment interaction by subgroup was significant. Reasonably large differences in clinician management decisions were observed among some scenarios with conflicting overall and subgroup findings. While there were no *right* answers, management decisions made by physicians with formal training in research methodology who had an academic appointment and were involved in research or who spent less time in patient care were less influenced by subgroup analyses. These physicians were more likely to focus on the overall results for management decisions.

Perceived Understanding: The Akl et al.[17] review found that among three included studies, data reported with natural frequencies (e.g., of 500 people treated, 350 experience improvement) had higher objectively measured understanding (pooled SMD 0.94, 95% CI 0.53 to 1.34) compared with data reported with percentages.[32] Using a repeated measures design, Cranney and Walley[35] found that among 73 general U.K. practitioners shown the same effectiveness data in four formats, only two practitioners identified that the formats presented the same data. Further, 75% reported having a hard time understanding statistics commonly found in journals. In a study previously described that assessed the effect of different presentations of meta-analysis data, Raina et al.[34] reported that only 12.5% of participants in the study were able to identify the correct definition for odds ratio, 25% for risk ratio, and 35% for risk difference.

Lastly, Froud et al.[40] conducted a qualitative research study among 14 purposively sampled U.K. clinicians to examine the clarity, ease, and perceptions of clinicians related to low back pain trial reporting methods using reports of five fictitious trials. Clinicians who had been previously involved in research tended to better recall statistical/epidemiological concepts and appeared more inclined to critically appraise the impact the trial had on their practice. Other themes identified in this study included participant concerns that individual RCTs were not sufficient forms of evidence to make an impact on practice and that current reporting methods are difficult to understand.

Discussion

Studies for Key Question 1 were mainly focused on evaluating physician knowledge and critical appraisal skills at a point in time or following educational/training interventions. Several studies reported attitudes, primarily confidence levels associated with interpreting statistics or medical research, and we identified only one study directly reporting on the relationship with prescribing behavior intentions. Of those studies reporting extant knowledge either through cross-sectional designs or as baseline measures prior to the intervention, the overall levels of knowledge and skills varied, but mean scores tended to be in the middle of the possible score range, at levels below what would likely be considered mastery. Although the evidence suggests that knowledge and skills can be improved with interventions such as journal clubs or workshops, we identified no studies that evaluated whether these improvements are durable over the long term or translate into changes in prescribing behavior. The evidence also suggests that physicians with additional formal training in biostatistics, epidemiology, or clinical research demonstrate higher levels of knowledge and appraisal skills than those with usual medical education.

Studies for Key Question 2 were focused on evaluating how different effect-measures used to report results influence physicians' perceptions of effectiveness and willingness to treat. The evidence suggests that the use of relative effect-measures increases perceptions of effectiveness and likelihood of treatment compared with absolute formats. Whether relative formats result in overestimates of effectiveness and overuse of treatment or result in underestimates of effectiveness and underuse of treatment cannot be discerned from these studies. Consistent with other reports,[41] these studies demonstrate that physician interpretation of data is influenced by the choice of effect-measure presented.

We note several limitations in this body of evidence and of this review. All studies were focused primarily on physician prescribers; no studies were specifically designed for nurse practitioner or physician assistant prescribers, though findings are probably applicable to other prescribers. We identified no studies that measured actual prescribing behaviors. A sizable proportion of the studies were conducted among convenience samples or had methodological issues that limited internal study validity. The use of idiosyncratic measures limited our ability to quantitatively synthesize findings across the body of evidence. Many Key Question 1 studies might be more appropriately categorized as curriculum evaluation and were not designed with the same rigor that might accompany intervention research. Indexing of these studies in electronic databases is inconsistent as evidenced by the proportion of eligible studies we identified through hand searches. Our review approach was also limited by the use of a single review to assess study quality and no formal assessment of publication bias.

A fundamental gap in this body of evidence is the lack of a common underlying framework as to how critical appraisal knowledge and skills fit among the myriad of influences on prescribing behavior. The path from a published clinical trial to an individual treatment decision by a prescriber is not linear and includes many contextual influences on decision making. This includes individual prescriber characteristics, patient values and preferences, system-level factors, professional community guidelines and standards, exposure to medical product promotion, and disease-specific considerations. For example, Brookhart et al.[42] evaluated therapeutic decision making related to osteoporosis and found that individual physician factors explained only 14% of the variation in decision making, and more than half of this could be attributed to clinic-level factors. In a synthesis of 624 qualitative studies, Cullinan et al.[43] identified four factors related to inappropriate prescribing in older patients: 1) the need to please the patient, 2) feeling of being forced to prescribe, 3) tension between prescribing experience and prescribing guidelines, and 4) prescriber fear. With the myriad of information sources available to prescribers, experts have suggested that teaching critical appraisal of primary research studies may be less important than teaching information management skills.[44]

Future research in this area should be framed within the larger context of information management, clinical decision support, and its relationship to safe and appropriate prescribing behaviors. This could include research that helps elucidate how information acquisition and understanding related to medications may be different within different contexts (e.g., point–of- care decision making versus formal continuing medical education) or for different kinds of prescribers. The use of contemporary behavioral theory and

implementation science frameworks to guide research, education, and policy in this area may offer a wider menu of options beyond critical appraisal to address the issue of safe and appropriate prescribing.

In conclusion, we identified 29 studies to address two key questions related to prescriber knowledge and critical appraisal skills of medical research and presentation formats that influence perceptions of effectiveness and intent to prescribe. Findings suggest that extant knowledge and skills are highly varied, and perhaps limited, and that certain features regarding how data are presented, such as relative versus absolute effect-measures, influence physician perceptions of treatment effectiveness and their intent to prescribe in hypothetical scenarios. No studies evaluated the influence of physician understanding or skills on actual prescribing behavior. Future research could be framed within the larger context of information management, clinical decision support, and its relationship to safe and appropriate prescribing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References

1. Lubloy A Factors affecting the uptake of new medicines: a systematic literature review. BMC Health Serv Res. 2014;14:469. [PubMed: 25331607]

2. Maggio LA, Tannery NH, Chen HC, ten Cate O, O'Brien B. Evidence-based medicine training in undergraduate medical education: a review and critique of the literature published 2006–2011. Acad Med. 2013;88(7):1022–1028. [PubMed: 23702528]

3. Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. Cochrane Database Syst Rev. 2001(3):CD001270. [PubMed: 11686986]

4. Association of American Medical Colleges. Physician Competency Reference Set (PCRS); 7 2013. https://www.aamc.org/initiatives/cir/about/348808/aboutpcrs.html

5. Young A, Chaudhry HJ, Pei X, Halbesleben K, Polk DH, Dugan M. A census of actively licenses physicians in the United States, 2014. JMR. 2015;101(2):8–23. https://www.fsmb.org/Media/Default/PDF/Census/2014census.pdf.

6. Ghosh AK, Ghosh K. Translating evidence-based information into effective risk communication: current challenges and opportunities. J Lab Clin Med. 2005;145(4):171–180. [PubMed: 15962835]

7. Harewood GC, Hendrick LM. Prospective, controlled assessment of the impact of formal evidence-based medicine teaching workshop on ability to appraise the medical literature. Ir J Med Sci. 2010;179(1):91–94. [PubMed: 19707728]

8. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, 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(7376):1338–1341. [PubMed: 12468485]

9. Lai NM, Teng CL. Self-perceived competence correlates poorly with objectively measured competence in evidence based medicine among medical students. BMC Med Educ. 2011;11:25. [PubMed: 21619672]

10. West CP, Ficalora RD. Clinician attitudes toward biostatistics. Mayo Clin Proc. 2007;82(8):939–943. [PubMed: 17673062]

11. Hadley JA, Wall D, Khan KS. Learning needs analysis to guide teaching evidence-based medicine: knowledge and beliefs amongst trainees from various specialities. BMC Med Educ. 2007;7:11. [PubMed: 17493274]

12. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. BMJ. 2004;329(7473):1017. [PubMed: 15514348]

13. Harris J, Kearley K, Heneghan C, Meats E, Roberts N, Perera R, Kearley-Shiers K. Are journal clubs effective in supporting evidence-based decision making? A systematic review. BEME Guide No. 16. Med Teach. 2011;33(1):9–23. [PubMed: 21182379]

14. Feldstein DA, Maenner MJ, Srisurichan R, Roach MA, Vogelman BS. Evidence-based medicine training during residency: a randomized controlled trial of efficacy. BMC Med Educ. 2010;10:59. [PubMed: 20807453]

15. Macrae HM, Regehr G, McKenzie M, Henteleff H, Taylor M, Fitzgerald GW, et al. Teaching practicing surgeons critical appraisal skills with an Internet-based journal club: a randomized, controlled trial. Surgery. 2004;136(3):641–646. [PubMed: 15349113]

16. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial [ISRCTN46272378]. BMC Med Educ. 2004;4(1):30. [PubMed: 15585061]

17. Akl EA, Izuchukwu IS, El-Dika S, Fritsche L, Kunz R, Schunemann HJ. Integrating an evidence-based medicine rotation into an internal medicine residency program. Acad Med. 2004;79(9):897–904. [PubMed: 15326018]

18. Thomas KG, Thomas MR, York EB, Dupras DM, Schultz HJ, Kolars JC. Teaching evidence-based medicine to internal medicine residents: the efficacy of conferences versus small-group discussion. Teach Learn Med. 2005;17(2):130–135. [PubMed: 15833722]

19. Cheatham ML. A structured curriculum for improved resident education in statistics. Am Surg. 2000;66(6):585–588. [PubMed: 10888136]

20. Dinkevich E, Markinson A, Ahsan S, Lawrence B. Effect of a brief intervention on evidence-based medicine skills of pediatric residents. BMC Med Educ. 2006;6:1. [PubMed: 16403214]

21. Shaughnessy AF, Gupta PS, Erlich DR, Slawson DC. Ability of an information mastery curriculum to improve residents' skills and attitudes. Fam Med. 2012;44(4):259–264. [PubMed: 22481155]

22. Sprague S, Pozdniakova P, Kaempffer E, Saccone M, Schemitsch EH, Bhandari M. Principles and practice of clinical research course for surgeons: an evaluation of knowledge transfer and perceptions. Can J Surg. 2012;55(1):46–52. [PubMed: 22269302]

23. Lao WS, Puligandla P, Baird R. A pilot investigation of a Pediatric Surgery Journal Club. J Pediatr Surg. 2014;49(5):811–814. [PubMed: 24851776]

24. Anderson BL, Williams S, Schulkin J. Statistical literacy of obstetrics-gynecology residents. J Grad Med Educ. 2013;5(2):272–275. [PubMed: 24404272]

25. Beasley BW, Woolley DC. Evidence-based medicine knowledge, attitudes, and skills of community faculty. J Gen Intern Med. 2002;17(8):632–639. [PubMed: 12213145]

26. Mimiaga MJ, White JM, Krakower DS, Biello KB, Mayer KH. Suboptimal awareness and comprehension of published preexposure prophylaxis efficacy results among physicians in Massachusetts. AIDS Care. 2014;26(6):684–693. [PubMed: 24116985]

27. Susarla SM, Redett RJ. Plastic surgery residents' attitudes and understanding of biostatistics: a pilot study. J Surg Educ. 2014;71(4):574–579. [PubMed: 24835297]

28. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA. 2007;298(9):1010–1022. [PubMed: 17785646]
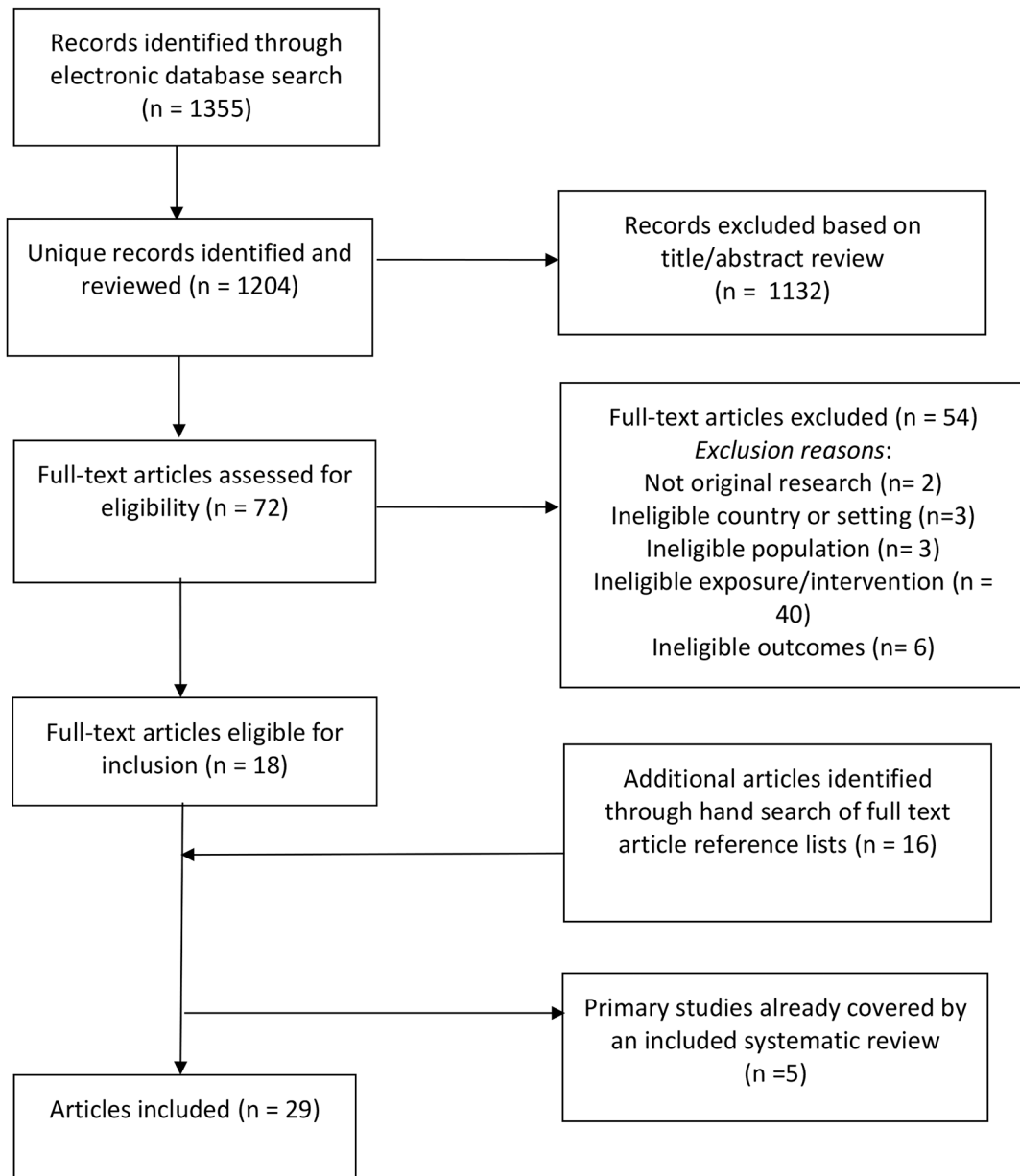
29. Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. J Med Educ. 1980;55(8):692–697. [PubMed: 7401147]

30. MacRae HM, Regehr G, Brenneman F, McKenzie M, McLeod RS. Assessment of critical appraisal skills. Am J Surg. 2004;187(1):120–123. [PubMed: 14706601]

31. Covey J A meta-analysis of the effects of presenting treatment benefits in different formats. Med Decis Making. 2007;27(5):638–654. [PubMed: 17873250]

32. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, Costiniuk C, Blank D, et al. Using alternative statistical formats for presenting risks and risk reductions. Cochrane Database Syst Rev. 2011(3):CD006776. [PubMed: 21412897]

33. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med. 1992;117(11):916–921. [PubMed: 1443954]

34. Raina PS, Brehaut JC, Platt RW, Klassen D, St. John P, Bryant D, Viola R, Pham B. The influence of display and statistical factors on the interpretation of metaanalysis results by physicians. Med Care. 2005;43(12):1242–1249. [PubMed: 16299436]

35. Cranney M, Walley T. Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. Br J Gen Pract. 1996;46(412):661–663. [PubMed: 8978112]

36. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med. 1992;92(2):121–124. [PubMed: 1543193]

37. Heller RF, Sandars JE, Patterson L, McElduff P. GPs' and physicians' interpretation of risks, benefits and diagnostic test results. Fam Pract. 2004;21(2):155–159. [PubMed: 15020384]

38. Lacy CR, Barone JA, Suh DC, et al. Impact of presentation of research results on likelihood of prescribing medications to patients with left ventricular dysfunction. Am J Cardiol. 2001(2):203–207.

39. Parker AB, Naylor CD. Interpretation of subgroup results in clinical trial publications: insights from a survey of medical specialists in Ontario, Canada. Am Heart J. 2006;151(3):580–588. [PubMed: 16504618]

40. Froud R, Underwood M, Carnes D, Eldridge S. Clinicians' perceptions of reporting methods for back pain trials: a qualitative study. Br J Gen Pract. 2012;62(596):e151–159. [PubMed: 22429424]

41. Fischhoff B, NT B, JS D, eds. Communicating Risks and Benefits: An Evidence-Based User's Guide. Silver Spring, MD: Food and Drug Administration (FDA), US Department of Health and Human Services; 2012.

42. Brookhart MA, Solomon DH, Wang P, Glynn RJ, Avorn J, Schneeweiss S. Explained variation in a model of therapeutic decision making is partitioned across patient, physician, and clinic factors. J Clin Epidemiol. 2006;59(1):18–25. [PubMed: 16360557]

43. Cullinan S, O'Mahony D, Fleming A, Byrne S. A meta-synthesis of potentially inappropriate prescribing in older patients. Drugs Aging. 2014;31(8):631–638. [PubMed: 24923385]

44. Slawson DC, Shaughnessy AF. Teaching evidence-based medicine: should we be teaching information management instead? Acad Med. 2005;80(7):685–689. [PubMed: 15980087]

Lessons for Practice

Physician prescribers may not have adequate critical appraisal knowledge and skills to understand and correctly interpret research evidence for safe and effective medication prescribing.

Knowledge and skills can be improved, at least in the near term, with a variety of education and/or training interventions.

Perceptions of treatment effectiveness and the intent to prescribe are increased by the use of relative effect-measures (e.g., risk ratio, odds ratios) in reports of research findings compared with the use of absolute effect-measures (e.g., risk difference), despite numerically equivalent effect sizes.

```
┌─────────────────────────┐
│ Records identified through │
│ electronic database search │
│      (n = 1355)           │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐          ┌──────────────────────────┐
│ Unique records identified and │────▶│ Records excluded based on │
│    reviewed (n = 1204)    │          │   title/abstract review   │
└─────────────────────────┘          │       (n = 1132)          │
            │                          └──────────────────────────┘
            ▼
┌─────────────────────────┐          ┌──────────────────────────────────┐
│ Full-text articles assessed for │──▶│ Full-text articles excluded (n = 54) │
│    eligibility (n = 72)   │          │     Exclusion reasons:             │
└─────────────────────────┘          │   Not original research (n= 2)     │
            │                          │ Ineligible country or setting (n=3) │
            │                          │   Ineligible population (n= 3)      │
            │                          │ Ineligible exposure/intervention (n = │
            │                          │            40)                     │
            │                          │   Ineligible outcomes (n= 6)       │
            ▼                          └──────────────────────────────────┘
┌─────────────────────────┐
│ Full-text articles eligible for │          ┌──────────────────────────┐
│    inclusion (n = 18)     │          │ Additional articles identified │
└─────────────────────────┘◀─────────│ through hand search of full text │
            │                          │ article reference lists (n = 16) │
            │                          └──────────────────────────┘
            │
            │                          ┌──────────────────────────┐
            │─────────────────────────▶│ Primary studies already covered by │
            │                          │ an included systematic review │
            ▼                          │       (n =5)              │
┌─────────────────────────┐          └──────────────────────────┘
│ Articles included (n = 29) │
└─────────────────────────┘
```

FIGURE 1. Disposition of studies (PRISMA Flow Diagram)
Preferred Reporting Items for Systematic Reviews and Meta-Analysis: Prescribers'
Knowledge and Skills for Interpreting Research Results

Biostatistical concepts	Epidemiological concepts	Evidence-based medicine concepts
Descriptive statistics (e.g., mean median, standard deviation)	Study designs	Formulating clinical questions
Hypothesis testing, type 1 errors, interpretation of p values and confidence intervals	Study biases	Searching for evidence
	Sub-group analyses	Critical appraisal of studies
	Study techniques to minimize bias (e.g., blinding)	• Determining internal validity
Power and type 2 errors	Measures of effect (RR, OR, RD, NNT, NNH)	• Determining applicability
Specific statistical tests (e.g., t-test, ANOVA, chi-square, Cox regression, Kaplan-Meier anlaysis)	Pre and post test probabilities	Applying evidence to patient care
	Sensitivity	
Meta-analysis	Specificity	
Types of variables	Likelihood ratios	
	Positive and negative predictive values	

Notes: ANOVA = Analysis of Variance, NNT = Number Needed to Treat, NNH = Number Needed to Harm, OR = Odds Ratio, RD = Risk Difference, RR = Risk Ratio

FIGURE 2.
Summary of Critical Appraisal Concepts Evaluated by Studies Included in Key Question 1

TABLE 1.

Extant and/or Post-intervention Knowledge and Critical Appraisal Skills among Physicians[a]

Author (Year) Study Design	Sample Description	Intervention/Exposure	Results
Akl et al. (2004)[17] Non-randomized trial	49 internal medicine residents	*Intervention:* 2-week rotation in EBM integrated into daily teaching rounds *Comparison:* no intervention	Berlin EBM Assessment [b] max score = 15 *Intervention:* Mean pre-intervention score (SD): 5.2 (2.9) Mean post-intervention score (SD): 6.4 (2.4) *Comparison:* Mean baseline score (SD): 8.3 (2.2) Mean follow-up score (SD): 6.8 (2.4) Adjusted mean difference in change between intervention and comparison group: 2.52 (95% CI 0.84 to 4.18, $p = .006$)
Anderson et al. (2013)[24] Cross-sectional	4,713 U.S. obstetrics and gynecology residents	Two-item knowledge assessment assessing statistical literacy	26% correctly answered the positive predictive value item 42% correctly answered the *p*-value item 12% correctly answered both questions
Beasley & Woolley (2002)[25] Cross-sectional	177 volunteer community-based statewide medical school faculty	EBM knowledge assessment	Mean % correct: 34 Self-reported understanding ("1" = not well" to "5" = extremely well) of various statistical and epidemiologic concepts ranged from 2.8 to 4.0
Cheatham (2000)[19] Repeated measures	12 surgical residents	Twelve-month curriculum in biostatistics and study design as part of monthly journal club	Mean % correct pre-curriculum: 59 Mean % correct post-curriculum: 79 $p < .004$
Dinkevich et al. (2006)[20] Repeated measures	69 pediatric residents	EBM training module for 2 hours a week for 4 weeks	Fresno EBM Test [c] Mean % correct pre-training: 17 Mean % correct post-training: 63 $p = .0001$
Lao et al. (2014)[23] Repeated measures	5 pediatric surgery fellows, residents, or attending physicians	Two-year long training curriculum in EBM, delivered over a six-session journal club format	EBM knowledge assessment, max score = 20 Mean score (SD) pre-training: 11.1 (5.7) Mean score (SD) early post-training: 16.2 (4.9) Mean score (SD) late post-training: 16.8 (3.6) $p = .05$ (comparison for pre vs. early post-session) $p = .03$ (comparison for pre vs. late post-session)
MacRae et al. (2004a)[15] RCT	81 general surgeons in academic and community-based practice settings	*Intervention:* virtual journal club (monthly mailed package of guided instruction for critical appraisal of two clinical articles) *Comparison:* clinical articles only	No baseline measures reported *Intervention:* mean % correct 58 *Comparison:* mean % correct 50 $p < .0001$
MacRae et al. (2004b)[30] Psychometric validation study	44 general surgery resident	*Intervention:* residents who were exposed to an epidemiologist led critical appraisal journal club *Comparison:* residents who were not exposed to the journal club	EBM assessment max score = 121 No baseline measures reported *Intervention:* mean score 56.6 *Comparison:* mean score 49.3 $p = .02$
Mimiaga et al. (2014)[26] Cross-sectional	115 HIV specialists and general physicians	Interpretation of data from two clinical trials of pre-exposure prophylaxis	72% correctly interpreted results of both trials 60% were able to specify the correct dosing of drug based on information presented in trials

Author (Year) Study Design	Sample Description	Intervention/Exposure	Results
Shaughnessy et al. (2012)[21] Repeated measures	22 family medicine residents	30 hours of face-to-face education in EBM concepts provided during residency orientation	Modified Fresno EBM test[c] max score = 160, passing score = 113 Mean score (range) pre-intervention: 104.0 (49–134) Mean score (range) post-intervention: 121.5 (65–154) $p = .001$ % with passing score pre-intervention: 40.1 % with passing score post-intervention: 73.4 $p = .025$
Sprague et al. (2012)[22] Repeated measures	62 surgeons, surgical residents, and medical students	A 2.5-day course for evidence-based surgery, study design, aspects of RCTs and statistics	Modified Fresno EBM Test[c], questions from CAMS test[d] and other assessment items. Mean % correct pre-course: 38.2 Mean % correct post-course: 51.7 $p < .001$
Susarla & Redett (2014)[27] Cross-sectional	22 interns, junior-and senior-level plastic surgery residents	Biostatistical knowledge assessment	Mean % correct (range): 53.0 (0 to 83.3)
Taylor et al. (2004)[16] RCT	145 health care practitioners	*Intervention*: single 3-hour EBM workshop *Control*: wait list	Knowledge assessment score range: −18 to +18; critical appraisal rated using 5 point Likert scale (1 = no ability, 5 = superior ability) No baseline measures reported *Intervention*: mean (SD) knowledge score 9.7 (5.3) Mean(SD) critical appraisal scores: methodology 2.4 (2.5), results 2.6 (28), generalizability 2.7 (2.2) *Control*: mean (SD) knowledge score 8.0 (5.1) Mean (SD) critical appraisal scores: methodology 2.0 (2.1), results 1.7 (1.8), generalizability 2.4 (1.7)
Thomas et al. (2005)[18] Non-randomized trial	46 internal medicine residents	*Interventions:* EBM conference group EBM small discussion group *Comparison*: no EBM exposure	EBM assessment, max score = 25 No baseline measures reported Mean score (SD) EBM small group discussion: 17.8 (4.5) Mean score (SD) EBM conference group: 12.2 (4.6) Mean score (SD) comparison group: 12.0 (4.5) $p = .014$ for small group v. conference group $p = .002$ small group v. comparison group
Weiss & Samet (1980)[29] Cross-sectional	141 internal medicine physicians	Epidemiology and biostatistics knowledge assessment	Max score = 10 Mean score (SD): 7.4 (1.6) Item-level mean % correct ranged from 40 to 97
Windish et al. (2007)[28] Cross-sectional	309 residents from 11 internal medicine residency programs	Knowledge assessment reflecting commonly used statistical methods and results in research literature	Mean % correct: 41.1 Item-level mean % correct ranged from 10.5 to 87.4 75% report not understanding all of the statistics encountered in the literature

[a] Study description and results excerpted from evidence tables provided in Appendix B of the Digital Supplement.

[b] Berlin Questionnaire: a 15-item test assessing EBM knowledge and skills in the five domains of assess, ask, acquire, appraise, and apply.

[c] Fresno Test: a 12 short-answer item test assessing ability to form an answerable question, list evidence-based resources, and interpret research results.

[d] Center for Applied Medical Statistics (CAMS) test: a test of statistical knowledge and application of statistics to real-world scenarios.

Notes: CI = confidence interval; EBM = evidence-based medicine; HIV = human immunodeficiency virus; PGY = post-graduate year; RCT = randomized clinical trial; SD = standard deviation