How good is the evidence to support primary care practice?

Mark H Ebell,[1] Randi Sokol,[2] Aaron Lee,[1] Christopher Simons,[3] Jessica Early[2]

10.1136/ebmed-2017-110704

[1]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia, USA
[2]Department of Family Medicine, Tufts University, Boston, Massachusetts, USA
[3]Tufts Family Medicine Program, Cambridge Health Alliance, Revere, Massachusetts, USA

Correspondence to:
Dr Mark H Ebell, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, 125 Miller Hall, UGA Health Sciences Campus, Athens, GA 30602, USA; ebell@uga.edu

Abstract

Our goal was to determine the extent to which recommendations for primary care practice are informed by high-quality research-based evidence, and the extent to which they are based on evidence of improved health outcomes (patient-oriented evidence). As a substrate for study, we used Essential Evidence, an online, evidence-based, medical reference for generalists. Each of the 721 chapters makes overall recommendations for practice that are graded A, B or C using the Strength of Recommendations Taxonomy (SORT). SORT A represents consistent and good quality patient-oriented evidence; SORT B is inconsistent or limited quality patient-oriented evidence and SORT C is expert opinion, usual practice or recommendations relying on surrogate or intermediate outcomes. Pairs of researchers abstracted the evidence ratings for each chapter in tandem, with discrepancies resolved by the lead author. Of 3251 overall recommendations, 18% were graded 'A', 34% were 'B' and 49% were 'C'. Clinical categories with the most 'A' recommendations were pregnancy and childbirth, cardiovascular, and psychiatric; those with the least were haematological, musculoskeletal and rheumatological, and poisoning and toxicity. 'A' level recommendations were most common for therapy and least common for diagnosis. Only 51% of recommendations are based on studies reporting patient-oriented outcomes, such as morbidity, mortality, quality of life or symptom reduction. In conclusion, approximately half of the recommendations for primary care practice are based on patient-oriented evidence, but only 18% are based on patient-oriented evidence from consistent, high-quality studies.

Summary box

What is already known about this subject?

- Clinicians are encouraged to base their practice on the best available evidence.
- However, much evidence is derived from subspecialty populations or addresses rare, uncommon problems.
- The extent to which recommendations for primary care practice are guided by high-quality evidence is unknown.

What are the new findings?

- In a primary care-oriented medical reference, 18% of recommendations were based on consistent, high-quality patient-oriented evidence (Strength of Recommendations Taxonomy (SORT) A), while approximately half were based on expert opinion, usual care or disease-oriented evidence (SORT C).
- Clinical categories with the most A recommendations were pregnancy and childbirth, cardiovascular, and psychiatric.
- Categories with the least were haematological, musculoskeletal and rheumatological, and poisoning and toxicity.

How might it impact clinical practice in the foreseeable future?

Primary care physicians should be aware that only a minority of recommendations are based on high-quality, patient-oriented evidence. This highlights the need for regular literature surveillance by primary care physicians to identify stronger evidence as it is developed. More research is needed in the primary care setting that evaluates the impact of interventions on patient-oriented health outcomes. Progress could be measured by periodically re-evaluating the percentage of recommendations based on high-quality evidence.

Introduction

Taking an evidence-based approach to practice means that physicians and other healthcare providers should make a conscientious effort to base clinical decisions on research-based evidence that is most likely to be free from bias, and integrated with clinical skills and patient values.[1] Evidence-based practice (EBP) has been an important paradigm shift in modern healthcare education and practice. Increasingly, medical references are taking a more rigorous, evidence-based approach to creating content, and some now systematically grade the strength of evidence for clinical recommendations.[2] 'Information Mastery', initially described by Slawson and Shaughnessy in 1994, adds a formal assessment of relevance to the assessment of validity, and suggests that physicians can reduce the work of finding evidence by relying on high-quality secondary sources of information. They emphasise that changes in practice should be based on research that is free of bias and also demonstrates improvements in patient-oriented outcomes, such as morbidity, mortality, quality of life and symptom improvement.[3 4]

Previous studies have examined the extent to which physician decisions are based on high-quality evidence. Evidence from randomised trials supported 14% of thoracic surgical treatments,[5] 48% of interventions in an inpatient palliative care unit,[6] 43% of ophthalmology interventions,[7] 65% of psychiatric interventions,[8] 24% of surgical interventions,[9] 59% of outpatient primary care[10] and 50%–53% of recommendations in medical inpatients.[11 12] In these

studies, the denominators are the decisions made by individual clinicians.

Essential Evidence Plus (http://www.essentialevidenceplus.com) is a comprehensive online medical reference published by Wiley-Blackwell. Its primary audience is primary care physicians, emergency physicians and hospitalists. Use of this reference has been shown to improve knowledge and attitudes towards EBP,[13–15] and, in combination with daily evidence summaries emailed to subscribers, leads to practice change and improvement.[16 17] Essential Evidence Plus includes 742 chapters, each focused on a different symptom, disease or procedure, as well as extensive collections of clinical calculators, decision support tools and critical appraisals of individual studies. It was designed around the principles of information mastery, with an emphasis on delivering valid, relevant, patient-oriented evidence. Each chapter uses the same organisational structure, based on a validated taxonomy of physician's questions at the point of care,[18 19] and includes from 2 to 12 'Overall Bottom-Line' recommendations for clinicians. These recommendations are graded 'A', 'B' or 'C' using the Strength of Recommendations Taxonomy (SORT), which integrates evaluation of relevance and validity (table 1). The SORT taxonomy was developed by the editors of family medicine journals and in addition to Essential Evidence is used by *American Family Physician* and the *Journal of Family Practice*, to widely read primary care journals in the USA.[20]

In the current study, we use the clinical recommendations made in Essential Evidence to better understand the extent to which recommendations for primary care practice are informed by high-quality research-based evidence. We also examine the extent to which these recommendations are based on relevant, patient-oriented evidence rather than expert opinion or disease-oriented evidence.

Methods

Methodology for evidence grading

The SORT system is used to grade each recommendation made in Essential Evidence as A, B or C (table 1). The strength of the SORT system is its simplicity, and the fact that it addresses both relevance and validity. Evidence is relevant if it is based on patient oriented rather than disease oriented or surrogate outcomes, and it is valid if the study was designed to avoid important biases and if studies are generally consistent in their findings. Level A and B recommendations must be based on evidence of improved patient-oriented outcomes, and level A recommendations additionally require that the evidence be consistent and free of important biases.

Each topic has 'Overall Bottom-Line' recommendations that reflect the most important recommendations regarding prevention, diagnosis, treatment and prognosis for each topic. The prevention (optional), diagnosis, treatment and prognosis sections each also have key recommendations graded using SORT.

The recommendations are written and graded by the author of each topic. The accuracy of each evidence grade is confirmed by an associate editor and by the editor-in-chief (MHE) during the editorial process. Each topic is reviewed every 4 months by an editor to add or update relevant Cochrane reviews and important original research, and is evaluated in detail every 1–2 years by the author and editor.

Topic classification

Topics were classified as belonging primarily to one organ system or specialty classification by the editor-in-chief of the publication. While each topic may be classified in two different categories within the reference, for example, pulmonary embolism may be considered both cardiovascular and respiratory, and postpartum depression both pregnancy and psychiatric, for the purpose of this analysis a single primary classification was selected.

Data abstraction

Essential Evidence has 742 chapters (also called topics). Of these, 21 do not include clinical recommendations, primarily because they describe how to perform a procedure, and were excluded. This left 721 topics for our review, which were divided into three groups of topics. Each group of approximately 240 topics was reviewed by two of the authors to abstract the category (overall, prevention, diagnosis, treatment and prognosis) and grade (A, B or C) of each recommendation for each topic. Reviewers reconciled any discrepancies (which were rare) by referring back to the chapter to reach consensus agreement.

Results

The 721 Essential Evidence topics included in our study made a total of 3251 'Overall Bottom-Line'

Table 1 Strength of Recommendations Taxonomy[20]

Strength of Recommendations	Definition
A	Recommendation based on consistent* and good quality† patient-oriented evidence‡
B	Recommendation based on inconsistent or limited quality patient-oriented evidence
C	Recommendation based on consensus, usual practice, opinion, disease-oriented evidence or case series

For a detailed discussion, see http://www.aafp.org/afp/2004/0201/p548.html.

*Consistency: most studies found similar or at least coherent conclusions, or high-quality and up-to-date systematic reviews exist and support the recommendation.

†Good quality: validated clinical decision rules, meta-analyses of high-quality studies and high-quality individual cohort studies for diagnosis; meta-analyses of RCTs, high-quality individual RCTs and all or none studies for treatment and prevention; and meta-analyses of good quality cohort studies and individual cohort studies with good follow-up for prognosis.

‡Patient-oriented evidence: outcomes that matter to patients, such as morbidity, mortality, symptom improvement, cost reduction and quality of life.

RCT, randomised controlled trial.

Table 2 Bottom-line recommendations by clinical category

	A	B	C	A or B*	B or C†	Total
Overall Bottom-Line	576 (18%)	1099 (34%)	1576 (49%)	1675 (51%)	2675 (82%)	3251
Prevention	192 (20%)	345 (37%)	403 (43%)	537 (57%)	748 (80%)	940
Diagnosis	74 (4%)	376 (19%)	1515 (77%)	450 (23%)	1891 (96%)	1965
Treatment	600 (26%)	890 (39%)	812 (35%)	1490 (65%)	1702 (74%)	2302
Prognosis	113 (7%)	894 (55%)	614 (38%)	1007 (62%)	1508 (93%)	1621

*Recommendations based on patient-oriented evidence (improvements in how well or long patients live).

†Studies with inconsistent evidence of benefit, low validity studies or based on disease-oriented evidence.

recommendations. These recommendations summarised the key points for clinicians related to prevention, diagnosis, treatment and prognosis. Thus, each chapter had a mean of 4.5 and a median of 4 'Overall Bottom-Line' recommendations for practice (range 2–12). There were an additional 6828 recommendations in the subsections on prevention, diagnosis, treatment and prognosis. The evidence grading is summarised in table 2.

Just over half (51%) of the 'Overall Bottom-Line' recommendations were based on patient-oriented evidence from original research (A or B). Recommendations regarding diagnosis were less likely to be based on patient-oriented evidence (A or B) than those for prevention, treatment or prognosis (23% vs 57%–65%). Because diagnostic accuracy alone is not considered a patient-oriented outcome, studies focusing on accuracy alone cannot be rated higher than C. A diagnostic study could be an A or B if a study demonstrated that use of the test improved a patient-oriented outcome.

Table 3 summarises the percentage of 'Overall Bottom-Line' recommendations that are based on patient-oriented evidence (A or B) for each organ system/specialty category. Five categories had at least 60% A or B recommendations: pregnancy and childbirth, cardiovascular,

Table 3 Overall bottom-line recommendations by organ system/specialty classification, sorted by the percentage of recommendations that are A or B

Category	Topics in category (n)	A	B	C	A or B*	B or C†	Total
Pregnancy and childbirth	16	26 (36%)	23 (32%)	23 (32%)	49 (68%)	46 (64%)	72
Cardiovascular	49	66 (30%)	85 (38%)	71 (32%)	151 (68%)	156 (70%)	222
Oncology	39	35 (20%)	78 (44%)	65 (37%)	113 (63%)	143 (80%)	178
Psychiatric	20	27 (28%)	33 (35%)	35 (37%)	60 (63%)	68 (72%)	95
Neurological	48	40 (19%)	85 (41%)	84 (40%)	125 (60%)	169 (81%)	209
Gastrointestinal	58	59 (20%)	112 (39%)	119 (41%)	171 (59%)	231 (80%)	290
Respiratory	36	36 (23%)	53 (34%)	69 (44%)	89 (56%)	122 (77%)	158
Congenital	23	16 (16%)	36 (37%)	46 (47%)	52 (53%)	82 (84%)	98
Gynaecological	29	27 (20%)	44 (32%)	66 (48%)	71 (52%)	110 (80%)	137
Skin disease	56	46 (19%)	82 (33%)	120 (48%)	128 (52%)	202 (81%)	248
Musculoskeletal and rheumatological	61	31 (11%)	102 (37%)	140 (51%)	133 (49%)	242 (89%)	273
Renal	20	13 (14%)	31 (33%)	49 (53%)	44 (47%)	80 (86%)	93
Male genitourinary	18	11 (15%)	24 (32%)	40 (53%)	35 (47%)	64 (85%)	75
Infectious disease	61	41 (15%)	77 (29%)	147 (55%)	118 (45%)	224 (85%)	265
Trauma and injuries	32	21 (15%)	39 (28%)	77 (56%)	60 (44%)	116 (85%)	137
Perinatal and infant	17	10 (12%)	26 (31%)	47 (57%)	36 (43%)	74 (88%)	83
Haematological	20	7 (8%)	31 (35%)	50 (57%)	38 (43%)	81 (92%)	88
Eye diseases	23	19 (18%)	25 (24%)	61 (58%)	44 (42%)	86 (82%)	105
Ear, nose and throat/allergy	36	17 (12%)	38 (28%)	83 (60%)	55 (40%)	121 (88%)	138
Endocrine	41	25 (13%)	47 (25%)	114 (61%)	72 (39%)	161 (87%)	186
Poisoning and toxicity	18	3 (4%)	24 (29%)	56 (67%)	27 (33%)	80 (96%)	83

*Recommendations based on patient-oriented evidence (improvements in how well or long patients live).

†Studies with inconsistent evidence of benefit, low-validity studies, or based on disease-oriented evidence.

oncology, psychiatric and neurology. Categories with the most C recommendations (ie, based on expert opinion or disease-oriented evidence) were poisoning and toxicity, eye diseases, ear, nose and throat/allergy, and endocrinology.

Level A recommendations are based on original research reporting consistent, high-quality, patient-oriented evidence. Only 18% of 'Overall Bottom-Line' recommendations were based on A level evidence. Level A recommendations were more common for treatment, not surprising given the greater emphasis on and funding for randomised trials of therapy than for trials comparing diagnostic strategies. They were most common in the clinical categories of pregnancy and childbirth (36%), cardiovascular (30%), psychiatric (28%) and respiratory (23%), and least common in poisoning and toxicity (4%), haematology (8%), and musculoskeletal and rheumatological (11%).

Discussion

Previous studies have used observation of physician decisions[6–12] to determine the extent to which physician decisions are based on high-quality evidence. In this study, we take a novel approach that evaluates the strength of evidence for a broad range of conditions in primary care and also assesses the extent to which that evidence is based on patient-oriented outcomes. This has the advantage of providing a more complete picture of the evidence supporting a wide range of clinical conditions, and has the potential to help us understand which fields have the greatest need for further research.

Few medical references rate the strength of evidence of all key clinical recommendations. One is Clinical Evidence, which reported that 11% of the treatments reported in randomised trials were beneficial and another 24% were likely to be beneficial (http://clinicalevidence.com/x/set/static/cms/efficacy-categorisations.html). We are not aware of any other comprehensive general medical reference that systematically rates the strength of evidence for all key clinical recommendations, or that systematically assess whether recommendations are based on patient-oriented evidence.

Our study used a different denominator than the previous observational studies of physician practice. While previous studies primarily examined recommendations made by physicians under observation in clinical practice, we considered all clinical recommendations for a broad range of symptoms and diseases. Nevertheless, we found that the percentage of recommendations based on good or fair quality patient-oriented evidence (51%) was similar to that seen in observational studies of medical inpatients and outpatient primary care (50%–59%).[10–12] A study of the 2711 recommendations in 14 cardiology guidelines found that 11% were level A evidence, 41% level B, and 48% level C, although these guidelines used slightly different grade definitions (A, multiple randomised trials or meta-analyses; B, single randomised trial or non-randomised trials; C, expert opinion, case studies or standards of care).[21]

Fields with an extensive tradition of performing randomised trials (cardiology, oncology and perinatology) had the highest percentage of A or B recommendations. Fields with the fewest A or B recommendations may have fewer clinical trials, or those trials may not measure patient-oriented outcomes, such as morbidity, mortality, symptoms or quality of life. For example, many studies of diabetes mellitus focus on glycaemic control, a surrogate outcome. These studies can be rated no higher than C using the SORT system, a decision that has been validated by recent studies showing lack of benefit and even harm from overly tight control of blood glucose.[22–24]

Limitations of the study include the fact that the reference is primarily designed for use by generalist physicians, emergency physicians and hospitalists; some highly specialised topics and diseases may be under-represented. However, this is likely to be a much more comprehensive set of recommendations than those observed in clinical practice, where physicians deal with a smaller subset of conditions on most days. The inter-rater reliability of the SORT system has not been formally assessed, but each rating is reviewed by the author, an assistant or associate editor, and the editor-in-chief (who developed the system and has used it for 9 years). It will be interesting to repeat this study in 10 years, to see if the percentage of recommendations based on high-quality, patient-oriented evidence is increasing.

Our findings point to a clear need for more research that evaluates improvement in patient-oriented outcomes, particularly in fields, such as endocrinology, ophthalmology, ear nose and throat/allergy, and haematology, as well as a need for more high-quality research on common primary care problems in general. Studies comparing diagnostic strategies and evaluating the effect of these strategies on patient-oriented outcomes are also lacking, and will be important as we assess the incremental value of diagnostic technologies. Examples of this type of study include a comparison of strategies using B-natriuretic peptide with those that did not have this information for the evaluation of dyspnoea,[25] and studies of the impact of clinical decision rules such as the Strep Score or Ottawa Ankle Rules.[26–28] However, they are relatively rare in the literature. Another potential area of study is evaluation of the inter-rater reliability of the SORT system, as well as other evidence classification schemes.

In conclusion, approximately half of the recommendations for primary care practice are based on patient-oriented evidence, but only 18% are based on patient-oriented evidence from consistent, high-quality studies.

Contributors The study was conceptualised and designed by Dr MHE. All of the authors participated in data extraction, reviewed drafts of the manuscript and approved the final work.

Competing interests MHE serves as editor-in-chief of Essential Evidence, for which he is paid by Wiley-Blackwell. He is also one of the co-developers of the SORT framework for grading evidence.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement We would be happy to collaborate with interested researchers and share our data with them as part of such a collaboration.

References

1. **Sackett DL**, Rosenberg WM, Gray JA, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;**312**:71–2.
2. **Prorok JC**, Iserman EC, Wilczynski NL, *et al.* The quality, breadth, and timeliness of content updating vary substantially for 10 online medical texts: an analytic survey. *J Clin Epidemiol* 2012;**65**:1289–95.
3. **Shaughnessy AF**, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *J Fam Pract* 1994;**39**:489–99.
4. **Slawson DC**, Shaughnessy AF. Teaching evidence-based medicine: should we be teaching information management instead? *Acad Med* 2005;**80**:685–9.
5. **Lee JS**, Urschel DM, Urschel JD. Is general thoracic surgical practice evidence based? *Ann Thorac Surg* 2000;**70**:429–31.
6. **Good P**, Stafford B. Inpatient palliative medicine is evidence based. *Palliat Med* 2001;**15**:493–8.
7. **Lai TY**, Wong VW, Leung GM. Is ophthalmology evidence based? A clinical audit of the emergency unit of a regional eye hospital. *Br J Ophthalmol* 2003;**87**:385–90.
8. **Geddes JR**, Game D, Jenkins NE, *et al.* What proportion of primary psychiatric interventions are based on evidence from randomised controlled trials? *Qual Health Care* 1996;**5**:215–7.
9. **Howes N**, Chagla L, Thorpe M, *et al.* Surgical practice is evidence based. *Br J Surg* 1997;**84**:1220–3.
10. **Lee HY**, Ahn HS, Jang JA, *et al.* Comparison of evidence-based therapeutic intervention between community- and hospital-based primary care clinics. *Int J Clin Pract* 2005;**59**:975–80.
11. **Nordin-Johansson A**, Asplund K. Randomized controlled trials and consensus as a basis for interventions in internal medicine. *J Intern Med* 2000;**247**:94–104.
12. **Ellis J**, Mulligan I, Rowe J, *et al.* Inpatient general medicine is evidence based. A-Team, Nuffield Department of clinical medicine. *Lancet* 1995;**346**:407–10.
13. **Leung GM**, Johnston JM, Tin KY, Kyk T, *et al.* Randomised controlled trial of clinical decision support tools to improve learning of evidence based medicine in medical students. *BMJ* 2003;**327**:1090.
14. **Grad RM**, Pluye P, Meng Y, *et al.* Assessing the impact of clinical information-retrieval technology in a family practice residency. *J Eval Clin Pract* 2005;**11**:576–86.
15. **Katerndahl DA**, Ferrer RL. Randomized trial of the effect of research design and publication characteristics on physician change. *Prim Care Companion J Clin Psychiatry* 2006;**8**:32–8.
16. **Grad RM**, Pluye P, Mercer J, *et al.* Impact of research-based synopses delivered as daily e-mail: a prospective observational study. *J Am Med Inform Assoc* 2008;**15**:240–5.
17. **Wang R**, Bartlett G, Grad R, *et al.* The cognitive impact of research synopses on physicians: a prospective observational analysis of evidence-based summaries sent by email. *Inform Prim Care* 2009;**17**:79–86.
18. **Ely JW**, Osheroff JA, Ebell MH, *et al.* Analysis of questions asked by family doctors regarding patient care. *BMJ* 1999;**319**:358–61.
19. **Ely JW**, Osheroff JA, Gorman PN, *et al.* A taxonomy of generic clinical questions: classification study. *BMJ* 2000;**321**:429–32.
20. **Ebell MH**, Siwek J, Weiss BD, *et al.* Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *Am Fam Physician* 2004;**69**:59–67.
21. **Tricoci P**, Allen JM, Kramer JM, *et al.* Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009;**301**:831–41.
22. **Duckworth W**, Abraira C, Moritz T, *et al.* Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med* 2009;**360**:129–39.
23. **Gerstein HC**, Miller ME, Byington RP, *et al.* Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008;**358**:2545–59.
24. **Patel A**, MacMahon S, Chalmers J, *et al.* Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2008;**358**:2560–72.
25. **Lam LL**, Cameron PA, Schneider HG, *et al.* Meta-analysis: effect of B-type natriuretic peptide testing on clinical outcomes in patients with acute dyspnea in the emergency setting. *Ann Intern Med* 2010;**153**:728–35.
26. **Stiell IG**, Wells GA, Hoag RH, *et al.* Implementation of the Ottawa knee rule for the use of radiography in acute knee injuries. *JAMA* 1997;**278**:2075–9.
27. **McIsaac WJ**, Goel V, To T, *et al.* The validity of a sore throat score in family practice. *CMAJ* 2000;**163**:811–5.
28. **Feldstein DA**, Hess R, McGinn T, *et al.* Design and implementation of electronic health record integrated clinical prediction rules (iCPR): a randomized trial in diverse primary care settings. *Implement Sci* 2017;**12**:37.