Chapter 23

Evidence-Based Decision-Making 1: Critical Appraisal

Laurie K. Twells

Abstract

This chapter provides an introduction to the concept of Evidence-based Medicine (EBM) including its history, rooted in Canada and its important role in modern medicine. The chapter both defines EBM and explains the process of conducting EBM. It includes a discussion of the hierarchy of evidence that exists with reference to common methods used to assess the levels of quality inherent in study designs. The focus of the chapter is on how to *critically appraise* the medical literature, as one step in the EBM process. Critical appraisal requires an understanding of the strengths and weaknesses of study design and how these in turn impact the validity and applicability of research findings. Strong critical appraisal skills are critical to evidence-based decision-making.

Key words Evidence-based medicine, Critical appraisal, Study design

1 Introduction

Evidence-based medicine (EBM) is "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" [1]. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research [1, 2]. More recently, it has been further defined as the integration of best research evidence with clinical expertise and *patient values* [3]. The process of EBM involves formulating a clinical question, searching and obtaining the best evidence to answer the question, critically appraising the evidence to ensure its validity and applicability, and implementing the findings in practice [1]. Dr. David Sackett is often regarded as the "father of evidence-based medicine" although the term is said to have been first used by Dr. Gordon Guyatt in the 1990s [4, 5]. EBM is a process that grew out of the need for medical education to move away from patient care based solely on "expert opinion" to that based on best evidence [3]. Although now just one step in the process, it is interesting that EBM grew out of critical appraisal—the assessment of the validity of scientific literature

Patrick S. Parfrey and Brendan J. Barrett (eds.), *Clinical Epidemiology: Practice and Methods*, Methods in Molecular Biology, vol. 1281, DOI 10.1007/978-1-4939-2428-8_23,

and its practical relevance to patient care [1]. Critical appraisal of the scientific literature was advanced by David Sackett and Brian Haynes at McMaster University in the early 1980s when they published a series of articles in the *Canadian Medical Association Journal* (CMAJ) entitled "How to read clinical journals" with various subtopics that included: the etiology or causation of disease, quality of care, the usefulness or harm associated with therapy, and the utility of diagnostic tests [6]. Following these articles, Sackett wrote the seminal text for students "Clinical Epidemiology: A Basic Science for Clinical Medicine" now in its 3rd edition and often referred to as "the bible" of EBM [3, 7, 8]. Over the next two decades, the *CMAJ* articles were further refined and led to the establishment of an EBM Working Group that subsequently developed a series of 25 papers known as the *JAMA User's Guide to the Medical Literature*. These guides were initially developed for clinicians to help them interpret the medical literature and support clinical decision-making [9]. The success of this series of papers provided the impetus for both the *JAMA User's Guide to the Medical Literature*, a textbook (in its 6th printing), as well as the development of a user-friendly, publically available website that houses numerous resources for supporting the practice of EBM (http://www.jamaevidence.com). The articles, text and website include a number of EBM resources, structured guides on how to appraise papers on topics such as therapy, diagnosis, prognosis, quality of care, economic analysis and overviews, and are considered by many as the definitive checklists for critical appraisal [10].

2 The Process of Evidence-Based Medicine

In the opening editorial of the very first issue of the journal *Evidence-Based Medicine*, the essential steps in this emerging science of EBM were summarized. These included: to convert information needs into answerable questions (i.e., to formulate the problem); to track down, with maximum efficiency the best evidence to answer these questions; *to appraise the evidence critically in order to assess its validity (or truthfulness) and its applicability (or usefulness)*; to implement the results of the appraisal into clinical practice and to evaluate performance [11, 12].

This process is often illustrated using Steps or an A's approach shown in Table 1. This chapter is an introduction to **Step 4**, to "*A*ppraise" the medical literature in order to assess its validity and applicability. The process of critical appraisal is a very important part, albeit one step, in the EBM process due to two key principles. First, not all evidence is considered equal, and second, a hierarchy of evidence exists linked to its design and inherent methodology.

Table 1
The process of evidence-based medicine

Step 1	*A*ssess important patient or policy problems
Step 2	*A*sk well-defined clinical questions from case scenarios, the answer to which will inform decision-making.
Step 3	*A*cquire information by selecting and searching the most appropriate resources
Step 4	*A*ppraise the medical literature for its validity (closeness to the truth) and its applicability (usefulness in clinical practice)
Step 5	*A*pply the results of the appraisal of medical literature to make sound, reasoned clinical decisions taking into account patient preferences and values
Step 6	*A*ssess or evaluate performance in applying the evidence

Appraising evidence requires an understanding of the strengths and weaknesses of epidemiological study design and how these in turn affect the validity and applicability of study findings [10].

3 Levels of Scientific Evidence

A number of classification systems have been developed to assess and describe the varying levels of evidence associated with different study designs. Although there is some debate over the strengths of individual study methods, there is a general consensus that a hierarchy of evidence exists. Various study designs will provide differing levels of evidence to support a treatment effect or causal relationship by limiting systematic bias [3, 8, 10]. This hierarchy of evidence is most often illustrated by a pyramid or similar graphic that places the types of evidence in the following order of decreasing strength:

1. Systematic reviews and Meta-analysis.
2. Randomized Controlled Trials.
3. Cohort studies.
4. Case–control studies.
5. Cross-sectional studies.
6. Case series/Case reports.
7. Expert opinion.

A very brief summary of these main study designs is provided here. For more detailed information please refer to other chapters in this textbook. Epidemiological research studies are divided into experimental/intervention or observational studies and with the exception of randomized controlled trials, the only

experimental study, most are observational in nature. At the top of the pyramid are studies that summarize other studies. Systematic reviews (SR) are produced by systematically searching, critically appraising, and synthesizing available literature on a specific topic (e.g., the difference between parental perception and actual weight status of children: a systematic review). A SR and meta-analysis includes a quantitative summary of all study results, the benefit being an increased power to assess the effectiveness (or lack of) of an intervention (e.g., the effectiveness and risks of bariatric surgery: an updated systematic review *and meta-analysis*). Clearly, the quality of the meta-analysis is dependent on the quality of the RCT's included. In some instances a high quality RCT will dominate the evidence base. In other instances, a meta-analysis will reveal a weak evidence base with few trials homogenous for the intervention, design, patient groups and outcomes. An RCT, considered the gold standard in study design, is the only study design whereby participants are randomly allocated to an intervention/experimental arm (e.g., new cancer treatment) or a control arm (e.g., standard of care + placebo). Follow-up takes place over time to measure one or more outcomes of interest. Within a cohort study, a group of individuals exposed to a risk factor (e.g., diabetes mellitus) is compared to a similar unexposed group and an outcome(s) (e.g., premature mortality) is assessed over a specific time period. Cohort studies can be either prospective or retrospective in nature depending on the nature of data collection. In a case–control study, a group of individuals with a disease/outcome of interest (e.g., birth limb defects) are identified and compared to a control group with respect to their past exposure status (e.g., medication use such as Thalidomide). Cross-sectional studies or prevalence studies classify subjects according to disease and exposure status. Data is often collected through health surveys and questionnaires (e.g., a health survey reports the prevalence of obesity and diabetes in a target population). A case report consists of a detailed report of a single patient while a case series provides information on more than one patient with the same features (e.g., four young men described with rare form of pneumonia, led to the discovery of AIDS) [7, 10, 11].

3.1 Methods Used to Evaluate Scientific Evidence

There are many examples of methods used by organizations to delineate the quality of evidence. Some of these include those developed by the: US Preventive Services Task Force (USPSTF); the Oxford Centre for Evidence-Based Medicine (CEBM) and the Grading of Recommendations Assessment, Development and Evaluation or GRADE working group.

3.1.1 The US Preventive Services Task Force

Varying levels of evidence are used to rank the effectiveness of treatments or screening tools relevant to the primary care environment and are classified using the following levels:

- Level I: Evidence obtained from at least one properly designed randomized controlled trial, well-conducted systematic review or meta-analysis of homogeneous RCTs.
- Level II-1: Evidence obtained from well-designed controlled trials without randomization.
- Level II-2: Evidence obtained from well-designed cohort or case–control analytic studies.
- Level II-3: Evidence obtained from multiple time series designs with or without the intervention; dramatic results from uncontrolled experiments.
- Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Prior to the grading of levels, individual studies are critically appraised for internal validity based on specific criteria unique to each study design. Ultimately each study will be described as good (if a study meets all criteria), fair (if a study does not meet one criterion but does not have a fatal flaw) or poor (the study has a fatal flaw) in terms of methodological quality. For example, when critically appraising an RCT the following descriptors could apply. A study could be described as (1) *Good:* if comparable groups were initially recruited and maintained throughout the study (follow-up at least 80 %); if reliable and valid measurement instruments were used and applied equally to the groups; if interventions were described clearly; if all important outcomes were reported, if confounders were taken into consideration and intention-to-treat (ITT) analysis was conducted. (2) *Fair*: although comparable groups were recruited at the start of the study period, questions in differences in follow-up exist; measurement instruments are acceptable and have been applied equally but may not be the best choice; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. ITT is conducted. (3) *Poor:* groups recruited at the start of the study are not close to being comparable or maintained throughout the study; unreliable/invalid measurement instruments are used or not applied consistently among groups (including not blinding outcome assessment); key confounders are not accounted for; and ITT analysis is absent. ((http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procmanual4.htm))

3.1.2 The Oxford Centre for Evidence-Based Medicine

The Oxford Centre for Evidence-Based Medicine (CEBM) provides a grading system (http://www.cebm.net/) to evaluate evidence for different types of questions that include those on therapy,

etiology, prevention, harm, prognosis, diagnosis, and economic analysis. The highest level of evidence is classified as 1a and refers to a SR with homogeneity (similar study methods) with the lowest level of evidence a 5 being expert opinion. An evaluation of evidence using these levels results in a recommendation by a grading system (A to D), with a Grade A recommendation suggesting consistent level 1 studies are available through to a Grade D recommendation that suggests only level 5 evidence is available or that alternate evidence is inconclusive.

3.1.3 The Grading of Recommendations Assessment, Development, and Evaluation

The GRADE working group (http://www.gradeworkinggroup.com) has further refined and developed the process of assessing the strength of a study by addressing more than just the quality of the research but also the impact other factors have on the confidence in study results. Similar to other systems, the quality of evidence is assessed on four levels (i.e., high, moderate, low, very low) while *the confidence factor* is based on judgments assigned in five different domains in a structured manner. For example, an RCT may be considered a high quality study with a low risk of bias, but depending on its assessment in other domains, it may be downgraded due to: risk of bias (e.g., no allocation concealment); imprecision (i.e., random error); indirectness (e.g., population, interventions or outcomes differ from those of interest). A body of evidence may be downgraded due to inconsistency (e.g., different point estimates with nonoverlapping confidence intervals) or publication bias (e.g., small sample sizes with large treatment effects, commercially funded research). Alternatively, an observational study of moderate quality could be upgraded due to a large effect size or evidence of a dose–response relationship and would further support inferences of a treatment effect.

4 Critical Appraisal: Basics

Critical appraisal is the process of systematically assessing the validity, usefulness, and relevance of the evidence [12]. The process can be divided into an examination of extrinsic and intrinsic factors. *Extrinsic factors* include taking note of the authors and their affiliations, the journal, the funder, and the stated conflicts of interest [13]. Examining the intrinsic factors requires a rigorous assessment of study design and methodology- the focus of critical appraisal. A number of excellent resources have been developed to support the critical appraisal process (see EBM Resources at end of chapter), and all use a very similar template that involves asking three main questions followed by a subset of specific questions associated with a particular type of question (e.g., therapy) or study design (e.g., cohort study). These questions include:

1. Are the results of the study valid?
2. What are the results?
3. Will the results help in caring for my patients? Are the results applicable or generalizable to my patient population?

For each main question, a number of publicly available EBM resources (e.g., JAMA User's Guides, Clinical Evidence-Based Medicine (CEBM, Cochrane Collaboration) provide checklists, templates, and worksheets to help health professionals and students learn how to effectively appraise the scientific literature in relation to its validity and applicability. In the section below, examples of the types of questions that should be addressed during the appraisal process are provided. This is not an inclusive list but an overview of the types of questions you would expect to answer when appraising an article. References throughout the chapter and in the reference section provide readers with some of the key resources that should be used in the process of critical appraisal.

I. Are the results of the study valid?
   The following questions are relevant for the appraisal of all research studies.
   i. Why is the research being conducted?
      a. Is a brief background or context provided as to why the study was conducted?
      b. What is the study about?
   ii. What is the research question being addressed?
      a. Is there a hypothesis being tested?
      b. Is the question described in a PICO format? (Population, Intervention, Control, Outcome)
      c. If, after I conduct a methodological assessment, the results are valid, are they applicable to my question, my patient or patient population? If yes, keep reading if no move to another paper.
   iii. What type of study has been conducted?
      a. Primary studies present original research, while secondary research summarizes or integrates primary research. A brief descriptor of the main types of studies and their objective is provided in Table 2.
   iv. Was the research study design appropriate to the type of question?

a. The clinical area and/or type of question will normally inform the appropriate choice of study design. Table 3 provides some examples to illustrate these choices.
For each type of study question and/or study design, a set of questions has been developed to assess the validity of the study methods. These questions help to assess whether selection biases (e.g., the groups being compared are different), or information biases (e.g., ascertainment of exposure status) exist as well as to determine the level of confounding that exists and how the authors have chosen to adjust for it.

v. Do the methods used increase the validity of the results? Broad questions for each study design include:

a. Systematic reviews and/meta-analysis—search details, comprehensivness and rigor of review, quality assessment, appropriate synthesis of results, heterogeneity

b. Randomized controlled trials—success of the randomization process (e.g., evidence of allocation concealment, equal groups), follow-up of patients, blinding, statistical analysis (e.g., ITT, per protocol), groups treated equally other than intervention

c. Cohort studies—recruitment of the cohort (e.g., is it representative of a defined population), the measurement of the exposure and outcome (e.g., subjective or objective measures), blinding (e.g., of the assessor), confounding (e.g., restriction, multivariate modelling, sensitivity analysis), loss to follow-up

d. Case–control studies—recruitment of cases (e.g., case definition, representative, prevalent vs. incident, sufficient sample size) and controls (e.g., representative, sufficient sample, matched), exposure ascertainment

e. Diagnostic studies—reference standard, disease status (e.g., level of severity), blinding.

II. What are the results?

i. What are the main results of the study? How are they presented? (e.g., Relative Risk, Odds Ratio, Hazard Ratio, % change, mean difference, sensitivity, specificity, likelihood ratios, Number Needed to Treat).

ii. Is the analysis appropriate to the study design?

iii. Are the results statistically significant? (e.g., *p*-values, Confidence Interval (CI))
iv. What is the treatment effect? Strength of effect?
    a. How precise is it? (e.g., width of CI)
v. Have the results been adjusted for confounding? (e.g., crude and adjusted analysis)
vi. Have drop-outs or lost to follow-up been accounted for? (e.g., ITT, per protocol analysis, sensitivity analysis)
vii. Do you believe the results? Could they be due to chance, bias or confounding?
viii. Do the results suggest a causal relationship?
    a. Guidelines have been developed to help assess the likelihood of a cause–effect relationship (see Assessing Causation below)
viii. Are you concerned about publication bias?

III. Are the results from the study applicable/relevant to my research question, patient or population of interest?
i. Can the results (or test) be applied to my patient/local population? (e.g., similar socio-demographic, health status, gender, age, country, health system)
    a. Are the results statistically significant and/or clinically significant?
ii. Were all relevant outcomes included in the study?
iii. Do the benefits outweigh the harms (if any)?

Table 2
Study design and its major objective

Study design	Major objective
Meta-analysis	To provide an overall summary statistic of multiple primary studies using an a priori protocol and integration of quantitative data from studies identified by a systematic review
Randomized Controlled Trial	To study the efficacy of a treatment or intervention
Cohort Study	To study prognosis, natural history of a disease or causation
Case–control Study	To identify potential causal factors for a disease or to study adverse effects
Cross-sectional Studies	To determine the prevalence of disease or risk factors

Table 3
The relationship between clinical area/type of question and research study

Clinical area	Type of question	Research study
Diagnosis	What disease is responsible for the abnormal findings?	Prospective, blind comparison to a gold standard Cross-sectional study
Therapy	What therapy is appropriate for a disease?	RCT Prospective cohort
Prognosis	What are the expected outcomes of a disease?	Longitudinal studies Retrospective/prospective cohort studies
Prevention	How can a disease be prevented or delayed?	RCT Cohort Case–control Case series
Harm	What intervention or other factor may be contributing to a disease?	RCT Cohort Case–control/Case series

4.1 Assessing Causation

Knowing what causes a disease or adverse outcome may be critical for understanding how to prevent, diagnose, treat or provide a prognosis. According to the Oxford Dictionary, a cause is defined as "something that gives rise to an action, phenomenon or condition" [14]. In the mid 1900s, Austin Bradford Hill and Richard Doll, who were responsible for the seminal studies on smoking and lung cancer, developed a guide to assess the causal relationship between an exposure and an outcome [8]. This is not a list of criteria or rules that have to be met, but a guide to help examine the strength of the available evidence in the context of a causal relationship between an exposure and an outcome (Table 4).

5 Concluding Remarks

The above types of questions and suggested resources will help to support critical appraisal of the scientific literature. These are the tools needed to systematically assess the validity, usefulness and relevance of available evidence. Evidence-based medicine has become synonymous with evidence-based health care or evidence-based practice. Critical appraisal is an important, albeit, one step in this process. Understanding the strengths, weaknesses and quality of study designs, and their inherent ability to provide high grade evidence for health interventions, will inform evidence-based decision-making and evidence-based practice.

Table 4
Austin Bradford Hill's guide for assessing causation [8]

Temporality	Exposure precedes disease
Experimental evidence	Evidence from true experiments
Strength	Exposure strongly associated with disease frequency
Biological gradient or dose–response	More exposure associated with higher disease frequency or severity
Consistency	The association is observed by different persons in different places during different circumstances
Coherence	The association is consistent with the natural history and epidemiology of the disease
Biologic plausibility	Causation is consistent with biological knowledge of the time
Specificity	One cause leads to one effect
Analogy	Cause and effect relationship has been established for a similar risk factor or disease

References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312:71–72
2. Sackett DL, Rennie D (1992) The science of the art of the clinical examination. JAMA 267(19):2650–2652
3. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (2000) Evidence-based medicine: how to practice and teach EBM, 2nd edn. Churchill Livingstone, Edinburgh
4. Smith R, Drummond R (2014) Evidence-based medicine-an oral history. JAMA 311(4):365–367
5. Guyatt G. (1991). Evidence-based medicine. *ACP J Club 114:*A-16.
6. Sackett DL (1981) How to read clinical journals, I: why to read them and how to start reading them critically. CMAJ 124(5):555–558
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) Clinical epidemiology—a basic science for clinical medicine, 2nd edn. Little, Brown and Company, Boston
8. Haynes RB, Sackett DL, Guyatt GH, Tugwell P (2006) Clinical epidemiology- how to do clinical practice research, 3rd edn. Lippincott Williams & Wilkins, Philadelphia
9. Guyatt GH, Rennie D (1993) Users' guides to the medical literature. JAMA 270(17): 2096–2097
10. Greenhalgh T (2001) How to read a paper. The basics of evidence-based medicine. BMJ Books, London
11. Fletcher RW, Fletcher SW (2005) Clinical epidemiology: the essentials, 4th edn. Lippincott Williams & Wilkins, Baltimore
12. Sackett DL, Haynes B (1995) On the need for evidence based medicine. Evidence-Based Medicine 1: 4–5
13. Booth A, Brice A (2004) Evidence-based practice for information professionals: a handbook. Facet Publishing, London
14. Oxford Dictionary. (2014). Oxford University Press. http://www.oxforddictionararies.com

Additional EBM Resources: Online Resources

JAMA User Guides. http://www.jamaevidence.com

Centre for Evidence-Based Medicine (CEBM) Oxford, UK. http://www.cebm.net/

Critical Appraisal Skills Programs (CASP) Oxford, UK. http://www.Students4bestevidence.net/

Greenhalgh T. (1997). How to Read a Paper. *BMJ* 315 (Series of ten articles)

The Cochrane Collaboration. http://www.cochranelibrary.com

CIHR KT learning modules. http://www.cihr-irsc.gc.ca/e/39128.html

KT clearing house (supported by CIHR and St. Michaels' Hospital and University of Toronto). http://ktclearinghouse.ca/cebm/

Evidence updates from the BMJ Evidence Centre: a collaboration between McMaster University and the BMJ Group. https://plus.mcmaster.ca/evidenceupdates/Default.aspx

Additional EBM Resources: Textbook Resources

Sackett DL, Straus S, Richardson S, Rosenberg W, Haynes RB (2000) Evidence-based medicine: how to practice and teach EBM, 2dth edn. Churchill Livingstone, London

Strauss S, Glasziou P, Scott Richardson W, Brian Hayes R (2011) Evidence-based medicine: How to practice and teach it, 4th edn. Churchill Livingstone, London

Guyatt G, Rennie D. eds. (2002). Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Chicago, IL: American Medical Association (3rd edition due out in September 2014])

Guyatt G, Rennie D (eds) (2002) Users guides: essentials of evidence-based clinical practice. American Medical Association, Chicago, IL

McKibbon A, Wilczynski N (2009) PDQ evidence-based principles and practice, 2nd edn. McGraw-Hill, Europe