Evidence to Guidelines : The Roadmap

Bobby Paul[1], Indranil Saha[2]

[1] Assistant Professor & Head of the Department of Preventive & Social Medicine, All India Institute of Hygiene & Public Health, Kolkata, West Bengal, India [2]

[2]Professor, Community Medicine, IQ City Medical College, Durgapur, West Bengal, India

Correspondence : Dr. Indranil Saha, E mail: drsahaindranil@gmail.com

Abstract :

Evidence based practice deals with current best medical evidence in conjunction with clinical expertise and patient values to guide health care decisions. This paper intends to draw attention of the readers to the brief history of Evidence Based Medicine and the different methods of evidence appraisal in quantitative research like "hierarchy of evidence" which lists a range of study designs ranked in order of decreasing internal validity and may vary according to research question. There are other methods like integrative methods, network meta-analysis, quantitative modeling and infectious disease modeling that are also used for evidence appraisal. There are various approaches of quality assessment of research evidence amongst which GRADE approach (Grading of Recommendations Assessment, Development and Evaluation) is discussed in a nutshell to outline the roadmap from evidence generation in research to guideline formulation. Fundamentally the GRADE approach provides guidance for rating quality of evidence and grading strength of recommendations in order to aid in evidence based health care decisions.

Keywords: Decision making, Evidence-Based Medicine/methods, Evidence-Based Medicine/ history, Policy making, Practice guidelines as topics, Public Health

Introduction :

Evidence based practice is the conscientious and judicious use of current best medical evidence in conjunction with clinical expertise and patient values to guide health care decisions and involves the rigorous process of tracking down the available research evidence, assessing its validity, and then using the "best" Evidence to medical decision making. The need for Evidence Based Medicine (EBM) is reiterated by the quote of Bernard Russell: "It is not what the man of science believes that distinguishes him, but how and why he believes it. His beliefs are tentative, not dogmatic; they are based on evidence, not on authority or intuition.[1] Now the formidable task is to navigate the evidence with critical evaluation skills and identify studies that should influence decision making and policy, keeping in minds the strengths and weaknesses of study designs in the medical literature i.e. internal validity, external validity, confounding, bias etc. The commonly used tool in this regard is the "hierarchy of evidence" or the levels of evidence, which lists a range of study designs ranked in order of decreasing internal validity. With this background, this paper intends to draw attention of the readers to the methods of evidence appraisal in quantitative research and the approaches of its quality assessment. Furthermore, the GRADE approach (Grades of Recommendation, Assessment, Development and Evaluation) which provides guidance for rating quality of evidence and grading strength of recommendations in health care is also explained in a nut shell in order to outline the roadmap from evidence generation in research to guideline formulation, an essential aid for clinical decision making.

Brief history of EBM :

The origin of EBM can be traced back to a report by the Canadian Task Force on the Periodic Health Examination in 1979. [2] The authors developed a

system of rating evidence generated from available research on the effectiveness of a particular intervention and further formulated a grading of recommendations based on the levels of evidence to prioritize the implementation of interventions in real world scenario. For example, Grade A recommendation was given if there was good evidence to support a recommendation that a condition can be included in the periodic health examination. This was subsequently adopted by the US Preventive Services Task Force and included methods for assessing the strength of evidence for public health decision making.

The foundation for evidence based practice was laid down by David Sackett, regarded by many as "the father of evidence based medicine" who defined it as a systematic approach to clinical problem solving which allows the integration of the best available research evidence with clinical expertise and patient values. [3] In the late 1960s, David Sackett, the professor of Medicine at McMaster University, in Ontario, Canada, along with his colleagues started teaching the students and internists the methods of critical appraisal of research as they believed that "A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system." [4] This was the beginning of a new era of treatment approach. Gordon Guyatt, the Director of the internal medicine residency program at McMaster University in 1990, further facilitated to usher in this era by asking the physicians to manage patients not on the basis of what authorities told them to do but on what the evidence showed worked, therein coining an appropriate terminology to this methodology as " Evidence Based Medicine." [5]

Evidence Appraisal in Quantitative Research

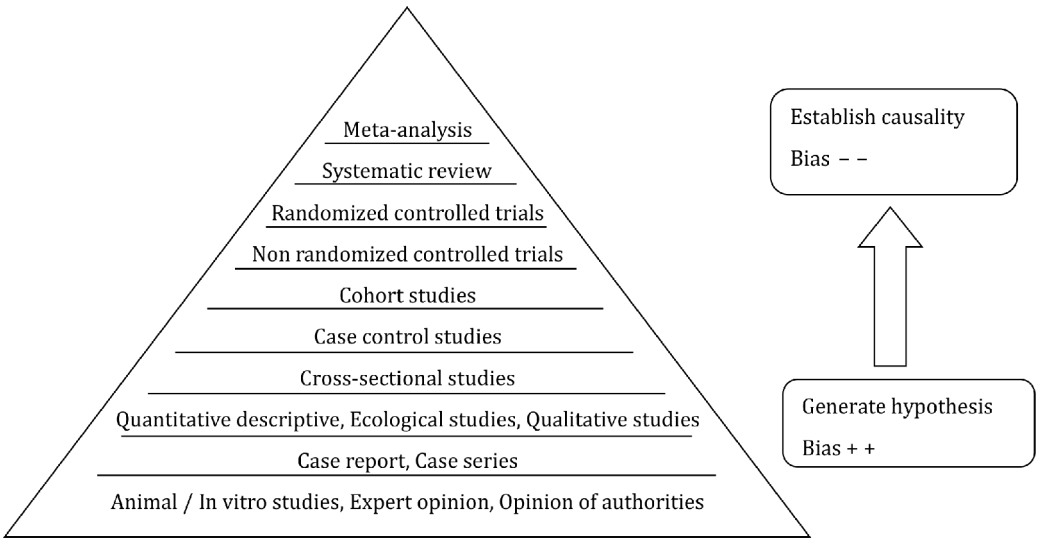
The hierarchy of evidence necessary for clinical decision making places higher value on study designs that focus on outcomes based on experiments and lower value on unsystematic clinical observation. Figure 1 depicts the hierarchy commonly used in quantitative research.[6] Although best suited for questions of therapeutic efficacy, it is either limited or no value for many other research questions such as appraisal of evidence for social or public health interventions, cost effectiveness of therapies etc. (Table 1).[7,8]

However there are other methods like Integrative Methods (Secondary or synthesis methods) that comprehensively consolidate findings of existing relevant research in order to resolve inconsistencies or ambiguities among existing studies and yield findings that may not have been apparent or significant in individual studies. [9] The findings from systematic literature review, meta-analysis, modeling (e.g., decision trees, state-transition models, infectious disease models), group judgment ("consensus development"), unstructured literature review and expert opinion may be combined or considered in a holistic manner to explore the broader social and economic contexts which will not only help in choosing the best treatment option for a disease, but also help to inform policies and guidelines, as the case may be, pertaining to the research question. [9]

Network meta-analysis (also known as multiple-treatment or mixed-treatment comparisons meta-analysis), which is an expansion of conventional pair-wise meta-analysis, is also currently used for development of clinical guidelines. This process analyzes simultaneously by both direct comparisons of interventions within randomized controlled trials (RCTs) and indirect comparisons across trials based on a common comparator (e.g. placebo or some standard treatment) when there are limited or no available direct ("head-to-head") trials of those interventions. [10]

Another important quantitative tool of EBM is quantitative modeling which is used to answer "What if?" questions i.e. the modeling techniques evaluate the clinical and economic effects of health care interventions. For example, decision analytic

Figure 1 : Hierarchy of evidence generation



modeling can be used to represent alternative sequences of clinical decisions for a given health problem. The probabilities are then calculated in terms of expected health outcomes and the cost effectiveness that would result from each strategy can also be computed. Decision models, often, are shown in the form of "decision trees" with branching steps and outcomes with their associated probabilities and values. [9]

Infectious disease modeling is yet another tool which is used to understand the spread of an infectious disease in a population with the help of a mathematical model. The disease in question is at first described in terms of transmission of the pathogen among hosts, depending on patterns of contacts among infectious and susceptible individuals, the latency period from being infected to becoming infectious, the duration of infectiousness, the extent of immunity acquired following infection, and other related factors. All the factors are then formulated in a complex mathematical model and the outcomes are used to make predictions about the number of individuals who are expected to be infected during an epidemic, the duration of the epidemic, the peak incidence, expected number of cases at each point in time and, indeed, the entire epidemic curve can be drawn. [11] It is especially useful to assess impact of control strategies and in situations when a randomized control trial is not possible because the disease of interest has not yet occurred in the specific population for which preventive/therapeutic strategies are to be formulated.

Assessment of Evidence Quality :

Quality of evidence is defined as the "extent to which one can be confident that an estimate of the effect or association is correct." [12] Many approaches have been used to assess the quality of a body of evidence since the 1970s. David Sackett emphasized on the importance of estimating types of errors and the power of studies when interpreting results from RCTs. For example, a poorly conducted RCT may report a negative result when in fact a real difference exists between treatment groups. [13] Different instruments are used to assess the reporting of different study designs which help to maintain the quality check in generating research evidence e.g. STROBE (Strengthening The Reporting Of Observational Studies), CONSORT (Consolidated Standards of Reporting Trials), PRISMA (Preferred Reporting Items of Systematic Reviews and Meta-Analyses Instrument), CHEERS (Consolidated Health Economic Evaluation Reporting Standards) etc.

However in recent years, there has been some convergence in these approaches, the efforts being put in this regard by many organizations such as the

Table 1: Hierarchy of best evidence according to research question

Research question	Hierarchy of best evidence
Effectiveness	RCT > Quasi -experimental > Analytic studies
Diagnosis	Studies of test accuracy among consecutive patients > studies of test accuracy among non consecutive patients > diagnostic case control studies
Prognosis	Inception cohort studies > studies of all or none > cohort studies
Economic evaluation	Decision model > Economic evaluation studies > expert opinion on incremental cost effectiveness of intervention and comparator
Meaningfulness	Qualitative or mixed method systematic review > Qualitative or mixed methods synthesis > Expert opinion
Safety	RCT > Quasi-experimental and analytic studies
Process of service delivery	Qualitative > Surveys > Evaluation studies (non experimental)
Acceptability of services	Qualitative > Experimental designs and surveys
Appropriateness of services	Qualitative research and cross section surveys
Systematic review of the above mentioned study designs are placed at the top of each hierarchy level	

Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, the Cochrane Collaboration , the US Agency for Healthcare Research and Quality Evidence-based Practice Centers (AHRQ EPCs) , the Oxford Centre for Evidence-Based Medicine , and the US Preventive Services Task Force (USPSTF). [9]

GRADE Approach for Guideline Development :

The GRADE approach (adopted by the Cochrane Collaboration, WHO and many others) provides the roadmap for grading the quality of evidence and develop and report recommendations to the guideline development group based on a common, transparent and sensible system. Fundamentally the GRADE approach is based on the philosophy of evidence based health care decisions that include the integrations of three domains namely 1) the health state (low or high income country) and circumstances that the patient presents with where decision making takes place; 2) the patient's populations or societal values and preferences how important are certain outcomes for decision making and 3) the actual underlying research evidence. These three domains are finally integrated to health care decision making.

Guidelines are recommendations intended to assist providers and recipients of health care and other stakeholders to make informed decisions. Recommendations are judgments based on quality of evidence, tradeoffs between benefits and harms, values and preferences of end-users, implementers and policymakers and with an implication on optimal resource use.

The guideline developers initially formulate the question which drive the evidence search usually in PICO (patient, intervention, comparator, outcome) format e.g. in babies born to HIV-positive women (P), does screening with a new rapid diagnostic test (I), compared with standard diagnostic methods (C) accurately detect disease (O)? [12] Then comes the task of evidence retrieval from published body of evidence for choosing the best possible outcome to be used for recommendation in the guideline. The outcomes retrieved from medical literature search are considered in terms of desirable (e.g. lower mortality, reduced hospital stay, reduced duration of disease, reduced resource expenditure etc) and undesirable effects (e.g. adverse reactions, development of resistance, costly treatment etc.) on patients and are then rated in order of importance by the guideline development group, the external review group and relevant stakeholders. The critically important outcomes, as chosen by the ratings are then further evaluated in GRADE. The starting point for rating the quality of evidence in GRADE is always the study design from which the outcome has been retrieved, which is broadly classified into two types:

Table 2 : Significance of the four levels of evidence

Quality	Definition	Implications
High	The guideline development group is very confident that the true effect lies close to that of the estimate of the effect	Further research is very unlikely to change confidence in the estimate of effect
Moderate	The guideline development group is moderately confident in the effect estimate: the true effect is likely to be close to the estimat e of the effect, but there is a possibility that it is substantially different	Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate
Low	Confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the true effect	Further research is very likely to have an important impact on confidence in the estimate of effect and is unlikely to change the estimate
Very low	The group has very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of the effect	Any estimate of effect is very uncertain

Table 3 : Determinants of strength of recommendation

Determinants	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation is warranted. Weak recommendation is warranted for smaller net benefit and lower certainty for that b enefit
Values and preferences	The greater the variability/ uncertainty in values and preferences, the more likely is a weak recommendation
Costs (resource allocation)	Strong recommendation is less likely to be warranted in case of higher costs of an intervention / higher resource consumption

- Randomized controlled trials (RCTs)
- Observational studies, including interrupted time-series (or quasi-experimental design), cohort studies and case-control studies, and other types of design such as case series and case reports.

Evidence based on randomized controlled trials is given a high-quality rating and evidence from observational studies is given a low-quality rating. These initial ratings are adjusted by the following factors, the presence of which can upgrade or lower the evidence.

The presence of factors that lowers the evidence quality are : [12]

1. Risk of bias
2. Inconsistency (heterogeneity),
3. Indirectness (lack of external validity)
4. Imprecision (when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect.)
5. Reporting bias

The factors that increase the evidence quality are: [12]

Magnitude of effect (no major threat to validity, consistent and direct evidence)

1. All plausible residual confounding may have reduced the demonstrated effect or increased the effect if no effect was observed
2. Large dose-response gradient

Once all outcomes that are critical for decision making have been evaluated, an evidence profile for each outcome is generated that provides estimates of the magnitude of desirable and undesirable consequences of an intervention and the confidence in those estimates to support a recommendation. Thereafter an overall GRADE of the quality of evidence is assigned for each outcome and categorized in four categories; high, moderate, low or very low (Table 2). [12]

This information is then provided back to the guideline panel who then formulate recommendations in a clear and unambiguous manner using

standardized wording, such as using the term "recommend" for strong recommendations and "suggest" for conditional or weak recommendations. The factors considered at this stage are summarized below (Table 3). [14]

Conclusion :

EBM tries to bring out the best practices from the best quality evidences. The roadmap from evidence generation through various study designs and the utilization of these evidences for guideline formulation is time consuming and complex. Nevertheless, the final outcome i.e. the evidence-based guideline is a scientific document, that categorizes the recommendations into strong and weak, is of utmost use to the evidence users i.e. the patients, clinicians, and policy makers. For patients, strong recommendation means that most people would want the recommended course of action; for clinicians / health care providers it means that most patients should receive the recommended course of action and for policy makers, the strong recommendation could be adapted as a policy in most situations and act as a guiding tool for planning, commissioning, and purchasing of healthcare services. [14] Finally the quality of guidelines can also be judged by using the AGREE II instrument (Appraisal of Guidelines Research and Evaluation) which judges the quality with the help of 23 items in 6 domains (scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, editorial independence) and two overall assessment domains. [15] Last but not the least, adaptation of trustworthy guidelines would not only improve patient outcomes by promoting beneficial interventions while discouraging those that are ineffective or possibly dangerous but also provide practitioners with credible guidance on appropriate, evidence based, and ethical practice and bring uniformity in patient management.

References :

1. Pai BH M, G Rajesh, Shenoy R. Research design hierarchy: Strength of evidence in evidence-based dentistry. Journal of Interdisciplinary Dentistry 2012; 2(3): 158–163.
2. Burns PB, Rohrich RJ, Chung KC. The Levels of Evidence and their role in Evidence-Based Medicine. Plast Reconstr Surg. 2011; 128(1): 305–310.
3. Evidence-based Medicine Working Group. Evidence-based medicine: A new approach to teaching the practice of medicine. JAMA 1992; 268(17): 2420-2425.
4. Glasziou P, Burls A, Gilbert R. Evidence based medicine and the medical curriculum. BMJ 2008; 337: a1253.
5. Smith R, Rennie D. Evidence-based medicine - an oral history. BMJ 2014; 348:g371.
6. Saha I, Paul B. Essentials of Biostatistics. Second Edition. Kolkata. Academic Publishers 2016. p. 168.
7. New JBI Levels of Evidence. Developed by the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party October 2013. Available from: joannabriggs.org/assets/docs/approach/JBI-Levels-of-evidence_2014.pdf. Last accessed on June 5, 2018.
8. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health 2003;57(7):527–529.
9. Goodman CS. HTA 101: Introduction to Health Technology Assessment. Bethesda, MD: National Library of Medicine (US); 2014. Available from: https://www. nlm.nih.gov/ nichsr/ hta101/ta10106.htm. Last accessed: June 5, 2018.
10. Kanters S, Ford N, Druyts E, Thorlund K, Mills EJ, Bansback N. Use of network meta-analysis in clinical guidelines. Bulletin of the World Health Organization 2016; 94(10): 782-784.
11. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. Clin Microbiol Infect. 2013; 19(11): 999–1005.
12. WHO handbook for guideline development. Second edition. World Health Organization. Geneva. 2014. Available from: http://apps.who.int/medicinedocs/documents/s22083en/s22083en.pdf. Last accessed June 5, 2018.
13. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1989; 95(2 suppl): 2S–4S. Schünemann HJ
14. , Oxman AD, Akl EA, Brozek JL, Montori VM, Heffner J et al. Moving from evidence to developing recommendations in guidelines: article 11 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proc Am Thorac Soc. 2012; 9(5): 282-92.
15. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. AGREE II: Advancing guideline development, reporting, and evaluation in health care. Preventive Medicine. 2010; 51(5): 421–424.