E–Health Systems Quality and Reliability:
Models and Standards

Anastasius Moumtzoglou
European Society for Quality in Healthcare, Greece

Anastasia Kastania
Athens University of Economics and Business, Greece

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 8

Quality Assurance in Evidence–Based Medicine

Ioannis Apostolakis
National School of Public Health, Greece

Periklis Valsamos
Greek Ministry of Health and Social Solidarity, Greece

Iraklis Varlamis
Harokopio University of Athens, Greece

ABSTRACT

Evidence-based medicine (EBM) refers to the careful examination of all the available evidence when making decisions about the care of the individual patient. It assumes that well known medical practices and solutions are combined with the patient's preferences and necessities in order to provide the most appropriate solution per case. The abundance of medical information in the web, the expansion of Semantic Web and the evolution of search services allowed the easier retrieval of scientific articles. Although the available infrastructure exists and continuously improves in performance, EBM still remains a complicated and sensitive process of high importance and has a need for Quality Assurance (QA). The purpose of this chapter is twofold: first, to provide an introduction on the concepts of Evidence-based Medicine, and second, to stress the necessity for structured methodologies that will assure the quality of the EBM process and ameliorate the final recommendations therapy. Since evidences are the building blocks of EBM, we capitalize on their quality and provide a critical overview of the existing methodologies in Quality Assurance of evidences.

INTRODUCTION

Evidence-Based medicine (EBM) can be thought of as the careful, explicit and reasonable use of patient related evidence (e.g. preferences, special needs etc.) in order to facilitate doctors in the selection of the most appropriate medical solution per case. It assumes the integration of individual clinical expertise with the best available external clinical evidence from systematic research (Sackett, 2003).

Tons of scientific journals, articles, patient guidelines and other related information are produced every day from scientific bodies and

research centers. The development of the Internet and other related technologies made sharing, distribution, searching and retrieval of scientific information easier than ever.

Clinicians can use the scientific databases available on the Internet (PUBMED, Medline etc) or the general purpose search engines in order to retrieve information quickly and effectively. Moreover, apart from these "pull services, modern tools (RSS, mailing lists etc) can "push" selected information to their subscribers. Also, the development of mobile technologies and wireless networks made the distribution of knowledge at the point of care easier than ever. A clinician can use a smart phone or a PDA to retrieve information at the point of care, or in other words "on the move".

Although significant progress is made in the area of the distribution, retrieval and searching of information, less is done in the area of its quality assurance. Consequently, the increase in information quantity had not an analogous impact in the quality of medical decisions. As a result the clinician is "left alone" to perform a time consuming, costly and error-prone process: the filtering and evaluation of the available information.

Truly, not the entire flood of provided knowledge is valid or useful for patient care. The study of Lundberg (Lundberg, 1992) on 100.000 scientific journals revealed that only 150 of those publications reported the 90% of all major scientific advances and less than 1,000 journals attained the 80% of the citations noted by Science Citation Index. The need to identify relevant information and to critically evaluate the scientific methodology and conclusions of the available information is obvious.

The purpose of this chapter is bifold. Initially, a short introduction in the concepts of evidence based medicine is given. This short introduction will provide the necessary definitions of EBM in order to avoid common misunderstandings and incorrect interpretations of the concept. Moreover, the importance of EBM for everyday clinical practice will be stressed.

In the following section we emphasize on the need for structured methodologies for the quality assurance and strength of recommendations. We focus on the problems that arise in the absence of a methodology, which assures the quality and relevance of provided information. Finally, we provide a critical review of existing methodologies in this field. The purpose of this presentation is to examine the proposed solutions for the quality assurance of the provided evidence as well as the provision of some suggestions.

BACKGROUND

Evidence-Based Medicine (EBM)

Clinicians in their everyday medical practice confront an overwhelming number of patients. In each medical session made, several questions arise concerning the proper prognosis, diagnosis and treatment. Moreover the differences between each individual patient case require the questions to be specialized according to the patient's medical condition, history and personal preferences. Truly, the selection of the "proper" treatment for each patient depends not only on scientific evidence but also from personal factors such as quality, personal beliefs and preferences of the patient.

Unfortunately, usually the decisions made by the clinicians are not supported by the suitable knowledge. The heavy workload and the absence of appropriate decision making tools hinder the clinicians from the careful processing of the available information and the selection of the most appropriate solution per incident. The lack of trustworthy and up to date information, make things even worse. As a result, the clinician is frequently left alone and her decisions are not adequately supported. Obviously, it is not practical for individual clinicians and patients to make these judgments unaided. In this context, Evidence Based Medicine (EBM) can be of great

help for the clinicians, providing the best possible evidence at the point of clinical care.

EBM employs scientific and engineering tools and techniques in order to collect medical evidence process them and apply results in medical practice. These tools comprise meta-analysis of medical literature, risk-benefit analysis, randomized controlled trials (RCTs) etc. EBM assesses the quality of evidence and evaluates the risks and benefits of various treatments. The Centre for Evidence Based Medicine defines EBM as the "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett, 2003). In other words evidence-based medicine integrates individual clinical expertise with the best available external clinical evidence from systematic research.

The algorithm for the practice of evidence-based medicine can be summarized in the following steps:

- **Define of the problem:** The first and most important step of the whole process. A misjudgment in the identification of the problem can disorient the clinician, lead to irrelevant questioning and consequently to wrong conclusions and actions. On the other side, the proper definition of the problem area can narrow the search space of the relevant literature and facilitate the clinician.

- **Select the appropriate clinical questions:** As previously, the accuracy of the clinical questions is crucial. An irrelevant or badly formed question will result to an unrelated or meaningless answer and a useless fact.

- **Track and appraise the best evidences:** Clinicians must search and utilize the most relevant available resources that answer the clinical questions, bearing in mind to gather both qualitative and quantitative evidences, which properly answer the ques-

tions. The quality of the selected information must be assessed, taking into account the validity of study results and their relevance to the questions.

- **Estimate the clinical importance of the evidence and the clinical applicability of any recommendation or conclusion:** In this step information is evaluated according to its relevance and applicability to the patient's problem. All evidences that are not applicable in a real clinical environment must be considered as a secondary source of information. Clinicians must compare the characteristics of their patient to those of the patients in the clinical study and verify that the study covered all important aspects of the patient's problem.

- **Integrate the evidence, the clinical expertise and the patient preferences and apply results to the clinical practice:** In cooperation with the patient, the clinician discusses the gathered evidence and suggested treatments.

- **Summarize and cache records for future reference:** An optional but highly recommended step in the process that maintains a "memory" for the system and builds useful knowledge base for the future.

Evidence based medicine is connected with some common misinterpretations, which should be carefully examined and avoided. A common misconception is that evidence-based medicine does not take into account clinical experience. This belief is wrong since signs and symptoms form the basis for the questions asked and guide the literature search. Moreover, evidence based medicine tries to support and back-up the clinician in her decision making process. Another wrong belief is that basic investigation and path physiology is not important for EBM. In contrast to this belief, the process of clinical problem solving followed by EBM has as a basic prerequisite a good

understanding of pathophysiology. Finally, the faulty perception that Evidence-based medicine ignores standard aspects of clinical training such as the physical examination is contradicted from the fact that Evidence-based practice considers the physical conditions of the patient while evaluating the evidence and also before applying treatment to the patient.

In conclusion, it is important to stress the factors that influence the adoption and implementation of the EBM (Freeman & Sweeney, 2001):

- **Clinician's experiences and personal beliefs**: The way the evidence is adopted and implemented is influenced by the doctor's experiences as well as her formed personal beliefs.
- **Clinician-Patient relationship:** The way the evidence is implemented is largely affected by the developed doctor-patient relationship, as well as by the specialized characteristics and beliefs of the patient.
- **Level of care provision:** The attitude towards evidence based medicine is different between clinicians of primary and secondary care.
- **Emotional-Psychological Factors:** EBM is not a "pure" intellectual process from which knowledge from studies and medical journals is transferred to the clinical practice. It involves also and an emotional part for both doctors and patients. Doctor's sometimes feel anxiety for the appliance of the new evidence or on the contrary may be neglecting it if the patient seems unwilling to follow new kinds of treatments. On the other hand, patients sometimes "jump" to new evidence and feel anxious to use them.
- **Developed Habits:** In some occasions, both patients and doctors are unwilling to follow new treatments and medications.

Quality Assurance of Medical Evidence

As already stated, the most important factor in the practice of EBM is to assure the quality of the evidence employed for supporting the medical decisions. In this context, it is crucial for the clinicians to follow a structured methodology and with the use of scientific tools to be able to evaluate the quality of evidence. In other words, it is necessary to reach the highest level of objectiveness during the evaluation of the validity, suitability, appropriateness, of the evidence used.

Before examining the methods for Quality Assurance of EBM, it is necessary to present the most widely employed techniques in EBM. These techniques originate from science, engineering and statistics and their results are the evidence for the Evidence based Medicine.

Randomized controlled trials (RCTs) are scientific experiments that evaluate the effectiveness of healthcare services and technologies over a population sample. The process assumes that a different solution is selected randomly from the set of available solutions and is applied to each test subject, thus eliminating causality and bias. RCTs can be *open*, *blind* or *double-blind*, depending on what degree the patient and the doctor are aware of the treatment. In an open trial the patient knows the full details of the treatment and a placebo effect is possible. Similarly in a blind process, it is possible that the treatment aware clinician can give hints to the patient about important treatment-related details, thus influencing the objectiveness of the study. On the contrary, in a double-blind trial the clinician is not informed on the treatment selected per case and as a result she is unable to affect the patient. In all cases, the administrator of the experiment is aware of all treatment allocations to patient-doctor pairs and thus is responsible to integrate the clinical results. The advantage of randomized controlled trials is that patients are not examined in isola-

tion but rather in groups (*control groups*). In this way, it is possible to compare results of different treatment methods, to evaluate them in reference to the group characteristics and to get valuable knowledge. The knowledge drawn from RCTs comprises clues on the effectiveness of the treatment, its side effects, the parameters that affect the treatment performance and is valuable for the decision making problem of treatment selection.

The randomness of assignments and the continuous monitoring of the trial results are necessary in order to avoid skewing of the results over the population groups. More specifically, a *randomization procedure* will generate a random and unpredictable sequence of allocations to patient groups at equal probabilities and the *allocation concealment* will guarantee that the group assignment of patients will not be revealed to the study investigators prior to definitively allocating them to their respective groups.

In **risk-benefit analysis** the risk of a decision and its expected benefits are put in the balance. In the sensitive case of patient care, the investigator must assure that the amount of benefit clearly outweighs the amount of risk. Those studies that have a clearly favorable risk-benefit ratio and guarantee not to harm the patient, may be considered ethical.

Meta-analysis is performed on the results of several studies that address a set of related research hypotheses. The results of the different studies are first aligned and then combined thus creating a fictional research output on a larger sample. The aggregated results have extended coverage and control and offer more powerful estimates of the true effect size than those derived in a single study under a given single set of assumptions and conditions. The meta-analysis on a group of studies can allow more accurate data analysis.

Clinical trials evaluate the safety and efficacy of new drugs, remedies or medical devices. They refer to products that have been already tested for quality and non-clinical safety. They comprise small scale pilot studies in the first step followed by larger scale studies when the initial results are positive in terms of safety and efficacy. They can vary in size from a single center in one country to multicenter trials in multiple countries. The medical products in a clinical trial are evaluated either individually or in comparison to existing products and the currently prescribed treatment.

Case-control studies originate from epidemiology and aim in locating the factors that affect a medical condition. They are performed on a set of subjects with similar properties and compare the *positive* (i.e. cases) and *negative* subjects (i.e. controls) in order to identify the minor differences that may affect their difference in condition. Case control studies examine the history of subjects and in order to locate past exposure to suspect factors that reasons their current condition. Although, they are easily applicable and require limited resources, they lack of large scale design and randomness. As a consequence, their conclusions are useful but cannot be widely applied in all medical cases. However, they can be employed in a preprocessing step for quickly and inexpensively identifying risk factors, and can be followed by a more profound analysis with more "credible" and comprehensive studies (e.g. randomized controlled studies). Moreover, they can be repeated over different population samples and fed as an input to a meta-analysis process.

Cohort or **panel studies** are widely employed in social sciences. They examine groups of people who are linked in some way, have experienced the same significant life event or share a common characteristic within a defined period in the past (e.g. birth, disease, leave school, lose their job, exposure to a drug, etc.) and compare their current behavior in a subject of interest (e.g. smoking). The analysis of cohort related information can tell us what circumstances in early life are associated with the population's characteristics in later life and allow us to find what encourages the development in particular directions and what can impede it. Similarly in medicine, a cohort study attempts to uncover the suspected association between cause

and disease; the negation of a hypothesis is refuted thus strengthening the confidence in the initial hypothesis. For this reason, the cohort should be identified and monitored a long time before and at least a short period after the appearance of the disease under investigation. The frequency of disease incidents, their severity, the geographical and temporal dispersion are some of the facts that should be evaluated in the results of a cohort study. Conducting a cohort study has a significant cost in time, people and money and thus it is a technique that should be utilized sparingly. Moreover, cohort studies are sensitive to erosion and are take a lot of time in order to generate useful data. Nevertheless, long-term cohort studies produce results of high quality, substantially superior to those of other techniques and are considered the "gold standard" in observational epidemiology. Less expensive research techniques can be employed to prepare the ground for a cohort study and further experimental trials can be utilized to maintain validity of the conclusions.

Cross-sectional studies refer to the concurrent observation of population subsets that expose significant differences in independent variables, such as IQ and memory. The subjects belong to different age groups and are examined at a single point in time. Cross-sectional research takes a 'slice' of its target group and bases its overall finding on the views or behaviours of those targeted, assuming them to be typical of the whole group.

Apart from the aforementioned techniques, evidence can be based on expert opinions (consensus practice guideline), or on literature review and do not include a systematic search.

In order to maximize the profit from combining evidence from medical research in the process of medical decision, Evidence based Medicine requests that the quality of resulting evidence is assessed. For the evaluation of evidence, several characteristics are examined:

- **Disease and Patient-Oriented Outcomes:** Outcomes that reflect the patient health status (e.g. blood sugar, blood pressure etc) and relate to the quality of life of the patients (i.e. the help them live longer an/or better lives) are of higher interest.
- **Research Evidence:** Evidence that is provided from original research is valuable but should be considered with care.
- **Level of Evidence:** It is tightly connected to the validity and structure of the study that produced the evidence. It is used for the results of individual studies but applies well on evidence that stem from multiple studies.
- **Strength of a recommendation:** Indicates the extent to which the conformance to this recommendation will do more good than harm. The strength (or grade) of a recommendation is based on a body of evidence (usually more than one study). Thus, in order to determine the strength of recommendation we take into account: the study that produced the evidence, the type of outcomes measured, the consistency and coherence of the evidence and the expected benefits, harms and costs.

SYSTEMS FOR QUALITY ASSURANCE OF EVIDENCE

The efficiency of Evidence-based Medicine is strongly connected to the appropriateness and quality of medical evidence. The multitude of research techniques and the abundance of evidence make it impossible for a doctor to become aware of all the related evidence and moreover to evaluate each one of them. As a result several efforts have been made in order to standardize and automate the quality assurance process for medical evidence and several systems have been developed in order to justify the quality of evidence.

One of the first efforts was made in 1979 by the *Canadian Task Force on Periodic Health Examination (1979)* and resulted to a classification

of the evidence produced by the different research methods. More specifically, evidence supported by Randomized Controlled Trials (RCTs) was classified as Level I (good), evidence supported by cohort and case control studies was classified as Level II (fair), and finally, evidence provided by expert's opinion was classified as Level III (poor). Consequently, the strength of recommendation was directly associated with the level of evidence that supported it. For example, a recommendation that was supported by Level I evidence was classified as a "strong" recommendation. The main advantage of this early system was its simplicity, which made it easy to understand and apply. On the other hand, there were many drawbacks, such as that it was based on many implicit judgments about the quality of randomized controlled trials.

Several systems, which have been developed since then attempted to provide alternative classifications and rate the evidence strength. The most important approaches are presented in the following.

U.S. Preventive Services Task Force (USPSTF) System

The U.S. Preventive Services Task Force (USPSTF) was established in 1984 aiming to provide a systematic review of medical evidence. Based on the Canadian Task Force System they presented their own grading system for the quality evidence and the strength of recommendation. As far as it concerns the quality of evidence they rated separately the individual study and the body of evidence and the service as a whole.

The classification of individual studies contained three main levels and several sublevels. More specifically:

*   Level I referred to studies that contained at least one properly designed randomized controlled trial
*   Level II comprised well-designed studies from one or more research groups.

Controlled trials without randomization were at the top of this level; cohort or case-control analytic studies followed; multiple time series with or without intervention and exceptional results in uncontrolled trials concluded this level.
*   Level III comprised the opinions of expert committees, the statements of respected authorities which were based on clinical experience and descriptive studies.

For the assessment of the body of evidence three criteria have been proposed, which refer to: the internal validity, the external validity, and coherence (i.e. consistency among studies and with other supporting evidence).

Finally, a three-point scale has been employed for rating the overall quality of the evidence. The levels were:

*   **Good:** For consistent results from studies of high quality. Such evidence demonstrates high applicability, direct and clinically important positive effects on the population.
*   **Fair:** For evidence that demonstrates clinically important positive effects, but is limited by the number, quality or consistency of the individual studies. Such evidence can be easily generalized to routine practice.
*   **Poor:** For those results that do not demonstrate positive effects on health outcomes. Such evidence is based on a limited number of poorly designed studies, which lack of important health outcomes.

At last, the quality of a medical service based on evidence is depicted to the strength of the produced recommendation. For this reason, a 5-level rating scale has been suggested:

*   **"A"** was accredited to services, which are based on good evidence and will potential-

ly improve health outcomes. The output of these services must be provided to eligible patients since their benefits substantially prevail over harms.

- **"B"** was used for services that evidently improve health outcomes. Their benefits outweigh harms and thus it is advisable to be provided to eligible patients.
- **"C"** refers to those services that can improve some health outcomes, but whose balance of benefits and harms is too close and thus cannot be a general recommendation. The decision is left to each individual patient and should take her preferences in account.
- **"D"** applies to services that are not recommended for use in asymptomatic patients. Usually, evidence shows that these services are ineffective and their harms outweigh the benefits.
- **"I"** applies to the services for which we have insufficient evidence and we are unable to recommend for or against their use.

The main strength of the system presented by USPSTF is that it provides a clear and direct linkage between quality of evidence and strength of recommendation. Moreover, it is more complete than the classification of the Canadian Task Force on Periodic Health Examination, since it takes into account other elements of evidence apart from the study design and it weighs benefits and harms. However, it has several limitations, since it is not adaptable to prognostic/diagnostic questions and it cannot provide recommendations in the absence of good evidence. Finally, the assessments do not always adjust for the individual patient values.

Oxford Centre for Evidence-Based Medicine (OCEBM) System

The Oxford Centre for Evidence-based Medicine (OCEBM) developed another set of Levels of Evidence and Grades of Recommendation, which was based on the grading system provided by the Canadian Task Force on the Periodic Health Examination (Ball, Sackett, Phillips, Straus & Haynes, 1998). OCEBM defines four main axes namely "therapy/aetiology", "prognosis", "diagnosis" and "economic analysis", which correspond to the broad type of clinical question. Each axe is divided into 5 broad levels of evidence ranging from 1 (least potential bias) to 5 (most potential bias), which mainly take into account the quality of design of each specific study. Additional factors estimate the outcome assessment ("minus" in case of imprecise result) and clinical sensibility (e.g. "appropriate spectrum" of patients).

Based on the level of evidence, OCEBM defines grades of recommendation strength (or grade of recommendation), which intrinsically is a mapping of levels of evidence to grades as follows:

- **Grade A** corresponds to Evidence Level 1 and comprises studies supported by randomized controlled trials, cohort studies, clinical decision rules validated in different populations etc.
- **Grade B** corresponds to Evidence Levels 2 and 3 and comprises studies such as consistent retrospective cohort, exploratory cohort, ecological studies etc.
- **Grade C** is for case-series studies and maps to evidence Level 3.
- **Grade D** comprises evidence based on experts' opinion, physiology, bench research or first principles etc.

The main advantages of the OCEBM system are the detailed classification of studies according to the level of evidence and the horizontal partitioning in the four axes that relate to diagnosis, aetiology, prognosis and economic analysis. However, the high level of detail may seem difficult for inexperienced users to follow. The main disadvantage of the system is the way

the translation of levels of evidence to grades of recommendations is made. Thus, no assessment is given of the clinical importance of the outcomes. Moreover, no balancing of benefits and harms is given. Finally, no assessment of the applicability of the studies is given.

American College of Chest Physicians (ACCP) System

The Consensus Conferences on Antithrombotic Therapy of the American College of Chest Physicians (ACCP) has developed guidelines to help clinicians make antithrombotic treatment decisions in average patients (Guyatt et al., 2001).

The Levels (quality of evidence) in the ACCP system are categorized as follows:

- **Grade A:** Randomized controlled trials (RCTs) with consistent results.
- **Grade B:** Randomized trials with inconsistent results, or with major methodological weaknesses
- **Grade C:** observational studies and generalization from randomized trials in one group of patients to a different group

In the ACCP system, the strength of recommendations is directly connected to the level of evidence.). The uncertainty associated with this trade-off will determine the strength of recommendations. The strength of the recommendation is denoted first using Grade 1 for strong and Grade 2 for weak recommendations. In this context, if experts believe that benefits outweigh risks then they will make a Grade 1 (strong) recommendation. On the opposite, if they are less sure they will make a Grade 2 recommendation. The grade is followed by the letter which denotes the quality of the evidence level (A, B and C), thus creating the following possible categories: 1A, 1B, 1C, 2A, 2B, and 2C.

The main advantage of the ACCP system is its simplicity. By simply checking the numeric grade the clinician can easily see if it is either a strong or weak recommendation. On the other hand, evaluating disease prognosis is not practicable with this approach. Moreover, this approach has been used little outside the antithrombotic therapy area.

Scottish Intercollegiate Guidelines Network (SIGN) System

The Scottish Intercollegiate Guidelines Network (SIGN) established in 1993 to develop evidence-based clinical guidelines for the National Health Service in Scotland (Harbour & Miller, 2001). The guidelines produced cover a wide range of healthcare professionals and clinical areas.

The quality of evidence is assessed using levels of evidence categorized from 1++ (least likely to be biased) to 4 (most likely to be biased). Studies are evaluated using critical appraisal checklists. These checklists have been originally designed in the Method for Evaluating Research and Guidelines evidence (MERGE) and are completed by clinicians of different background and degrees of expertise. Based on the qualitative assessment of answers to this checklist, we are able to define the evidence's quality level. Different questions are used to appraise the different types of study.

The strength of recommendations is graded using a scale from A to D. The grade of recommendation is drawn from the level of evidence and clinical judgement. The later includes the size and consistency of the body of evidence, its applicability, clinical impact and generalisability.

The main advantages of the SIGN system are its simplicity, and its potential to discriminate between study design requirements for different clinical questions. On the other hand, there are disadvantages too. The grades of recommendation have and unstructured formation. The "considered judgement" has many areas to be considered. Assimilation of the other factors is not well de-

scribed. Finally, there is no way of assessing or challenging these considerations.

Australian National Health and Medical Research Council (ANHMRC) System

The Australian National Health and Medical Research Council (ANHMRC) provides a framework for evaluating the strength of evidence across several dimensions (National Health and Medical Research Council, 2000), which relate to the type, size and randomness of the study.

First, the level of evidence depicts the quality of the study design and the scale is as follows:

- Level I: Contains evidence obtained from a systematic combined review of all relevant randomized controlled trials
- Level II: Is limited into evidence obtained from at least one properly-designed randomized controlled trial
- Level III: Has three sublevels comprising evidence obtained from well-designed pseudorandomised controlled trials, comparative studies with concurrent controls and allocation not randomized and comparative studies with historical control respectively.
- Level IV: Contains evidence obtained from case series.

Second, the quality of the evidence is assessed using methods that measure the bias of the study and its effects on the results. For each study type, standard quality assessment methods have been developed.

Third, the statistical precision of the evidence is assessed. In this context, the magnitude of the P-value and the precision of the estimate of the treatment effect are important. Similarly, the *size* of the treatment effect (i.e. the distance from the null value) is assessed as an indicator of the evidence usability.

Last but not least, the system evaluates the relevance of evidence as a measure of appropriateness of the outcomes to the specific case. The importance for the patient, the duration of effects and the applicability of the study findings to different settings and patient groups are examined.

Each recommendation is accompanied with a checklist that summarises the data and classifies it according to the dimensions of evidence strength (level of evidence, quality of evidence, statistical precision,. relevance and size of treatment). The checklist summarizes the results from the synthesis of the available evidence. Opposite to other systems, there is no single strength of recommendation climax.

The main advantage of the ANHMRC system is the multidimensional evaluation of the strength of evidence. In this manner, it allows clinicians to focus on the dimensions that are more important to them and to combine more than one dimension by applying weights to each one of them, according to their interest. On the other hand, the absence of a single classification system for the strength of recommendations is one of the major drawbacks of the system. Moreover the system does not evaluate fully the applicability of the results to individual patients, but covers them in a separate guide. Finally benefits, harms and costs are not integrated in the process.

U.S. Task Force on Community Preventive Services (USTFCPS)

The *Guide to Community Preventive Services* (*Community Guide*) is being developed by the non-federal Task Force on Community Preventive Services (Task Force) and is supported by the U.S. Center for Disease Control and Prevention (CDC) and others (Truman et al., 2000).

USTFCPS provides systematic reviews and evidence-based recommendations that can be applied on population-level and not on single patients. Consequently, the evaluation of population-based interventions differs from that of

individually-oriented clinical care interventions, which was the subject of all the aforementioned systems. The systematic reviews are conducted by various teams of researchers. The effectiveness and quality of each individual study is assessed, the results are extracted and analyzed. To give an example, randomized controlled trial (RCT) is less critical in population-based research than it is in a clinical research. More specifically, it is not always ethical or feasible, may have limited internal validity, or may have serious threats to external validity.

The body of evidence is characterized as strong, sufficient or insufficient using as criteria the strength of their design and execution, the number of available studies and the size and consistency of reported results. The suitability of study design is based on characteristics that protect against potential threats to validity. Finally, the quality of the study is affected by the execution details, such as: the population of the study and the descriptions of intervention, the population sampling, the exposure and outcome measurement, the data analysis method employed, the interpretation of results (including follow-up, bias, and confounding), etc.

The result of this analysis is to characterize a study for having good, fair, or limited quality of execution. Decision is based on the number of limitations noted, with values close to 0 for good execution and 5 for limited execution performance. The latter studies are not used to support recommendations. Sufficient or strong evidence can be based either on a small number of studies with better execution and more suitable design or a larger number of studies with less suitable design or weaker execution.

Consistency of results is defined as being generally consistent in direction and size based on the opinion of the Task Force. *Effect sizes* are defined to be large, intermediate or small based on the opinion of the Task Force. In general, larger effect sizes (e.g., absolute or relative risks) are considered to represent stronger evidence of effectiveness than

smaller effects. *Expert opinion* can be applied by the Task Force when other evidence is not available. The strength of evidence is related directly to the strength of recommendations.

One of the main advantages of the Community Guide is that it allows the participation of people from different backgrounds and perspectives and thus helps minimizing institutional and individual bias. Moreover, it supports the decision making process with different kinds of evidence (e.g. effectiveness, economic evaluations, etc.) and takes into account many factors (e.g., study design, study execution, numbers of studies, etc.) when assessing the process effectiveness. Its main disadvantage is the high complexity. Furthermore, it is demanding in time, resources and expertise and strongly dependent on the Task Force opinions.

Grade Working Group System

The Grade Working Group takes in account more dimensions than just the quality of medical evidence. It performs data extrapolation, thus allowing research outcomes to be employed in situations that significantly differ from that of the original study. Thus, the quality of evidence used to support a clinical decision is a combination of the quality of research data and the clinical 'directness' of the data (Atkins, Best & Briss, 2004).

Although these systems have several differences their aim remains practical the same: Guide clinicians and users in the selection of the most valid and trustworthy evidence.

CRITICISM OF METHODOLOGIES

It is obvious that the lack of a concrete methodology for evidence quality assurance has several drawbacks and creates many problems. Firstly, there is no "common ground" for the evaluation of evidence quality, resulting in controversial opinions and complicating the decision process for the clinician. Secondly, without the justification

and documentation provided by these methodologies the usage of the evidence and its results may be wrong. Any scientific results apply to several conditions and populations which must be taken into account before its application. Also, the compliance with such structured methodologies provides a common ground-common language for the way in which evidence should be evaluated enabling knowledge transfer between the clinician and healthcare organizations.

The factors that can lower our confidence about the quality of evidence (Guyatt et al., 2008) comprise of:

- The limitations of the study such as lack of blinding, no report of outcomes etc.
- The incapability to justify the causes of variability.
- The indirect comparison of population samples, methodologies, therapies etc.
- The small size of the sample and the wide confidence intervals.
    On the contrary, several factors can increase our confidence about the quality of evidence:
- The magnitude of the effect and the absence of bias. Observational studies for example give low-quality evidence due to the large number of unknown parameters that could not be measured. However, they are based on larger population sample and are more resistant to bias. Thus, the evidence is stronger.
- If all plausible confounding would decrease the magnitude of effect, this increases the quality of the evidence, since we can be more confident that an effect is at least as large as the estimate and may be even larger.

Methodologies and tools used in EBM usually use statistical inference methods in order to generalize the study outcomes and make them applicable to a particular population/sample. The factors related with each individual patient are numerous and consequently the complexity and uncertainty in the projection of results is a great obstacle that should be handled with skepticism (Atkins, 2008). In many cases the knowledge retrieved from clinical research cannot answer the primary question of what is best for the particular case of the patient at hand. Although EBM is not meant to replace clinical practice and examination it can ideally act as a complement.

In a similar manner, the projection of studies results to different populations or time periods should remain in question. Also, the variations of the quality of studies complicate the generalization of results. Moreover, in some medical cases (such as surgeries) the randomized controlled trials can be considered as unethical. Also, historically, certain groups are been under-researched. This lack of available research affects the quality of evidence and does not allow the generalization of results (Rogers, 2004).

Randomized controlled trials are useful for examining therapies effectiveness for controlled medical conditions ("normal situations"), but for complex patient situations the effect of each treatment is difficult to be evaluated. As a result, some studies conclude in results of small significance. Moreover, RCTs give evidence of high quality but are rather expensive. Since research is strongly depended on the available funds, several areas of research receive less interest than others. For example pharmaceutical companies, traditionally fund studies that investigate the efficiency and safety of drugs, but this is not common for the majority of studies. Another obstacle is that not all studies are published and consequently are not accessible to everyone. This results in leaving important parts of the available evidence out of the literature, and makes them useless for EBM (Friedman & Richter, 2004). Finally, the results reported in a clinical trial or study may be higher

compared to those of the real clinical practice due to closer patient monitoring during trials.

CONCLUSION

The chapter provides an introduction to the main concepts of Evidence Based Medicine giving emphasis on the evaluation of the quality of evidence. The main aim of EBM is to provide a different paradigm in the everyday clinical practice, which defines documentation of medical decisions, justification and comparative analysis of evidence and leads to less error-prone and more qualitative medical treatment. Due to the high variability of the human factor, attention should be taken for the proper evaluation of the provided evidence. In this context, we have presented the main methodologies from literature that attempt to standardize the ways the evidence is evaluated. The details, advantages and disadvantages of each methodology have been detailed thus providing the ground for the compliance and homogenization of these systems in the future. Our next step is to examine the available semantic technologies such as ontologies and other knowledge representation models, which can be employed for proper defining the quality assessment systems and systematically incorporate them into the evidence based medicine paradigm.

REFERENCES

Atkins, D. (2008). *The limits of evidence based medicine*. Retrieved January 15, 2009, from http://www.dbskeptic.com/2008/08/17/the-limits-of-evidence-based-medicine/

Atkins, D., Best, D., & Briss, P. A. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, *328*(7454), 1490. doi:10.1136/bmj.328.7454.1490

Ball, C., Sackett, D., Phillips, B., Straus, S., & Haynes, B. (1998). *Levels of evidence and grades of recommendations*. Retrieved January 17, 2009, from http://www.cebm.net/levels_of_evidence.asp

Canadian Task Force. (1979). Canadian Task Force on the Periodic Health Examination: The periodic health examination. *Canadian Medical Association Journal*, *121*, 1193–1254.

Freeman, A. C., & Sweeney, K. (2001). Why General Practitioners Do Not Implement Evidence: Guidelines for Reading Literature Reviews. *British Medical Journal*, *323*(7321), 1100–1104. doi:10.1136/bmj.323.7321.1100

Friedman, L. S., & Richter, E. D. (2004). *Relationship between conflicts of interest and research results*. Retrieved January 18, 2009, from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=14748860&itool=iconabstr

Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schunemann, H. J.Grade Working Group. (2008). What is 'quality of evidence' and why is it important to clinicians? *British Medical Journal*, *336*(1), 995–998. doi:10.1136/bmj.39490.551019.BE

Guyatt, G. H., Schünemann, H., Cook, D., Pauker, S., Sinclair, J., Bucher, H., & Jaeschke, R. (2001). Grades of recommendation for antithrombotic agents. *Chest*, *119*(1), 3S–7S. doi:10.1378/chest.119.1_suppl.3S

Harbour, R., & Miller, J. (2001). A new system for grading recommendations in evidence based guidelines. *British Medical Journal*, *323*(1), 334–336. doi:10.1136/bmj.323.7308.334

Lundberg, G. D. (1992). Perspective from the editor of JAMA. *The JAMA Bulletin of the Medical Library Association*, *80*(2), 110–114.

National Health and Medical Research Council. (2000). *How to use the evidence: assessment and application of scientific evidence*. Retrieved January 18, 2009, from http://www.nhmrc.gov.au/publications/synopses/cp65syn.htm

Rogers, W. A. (2004). Evidence based medicine and justice: a framework for looking at the impact of EBM upon vulnerable or disadvantaged groups. *Journal of Medical Ethics*, *30*(1), 141–145. doi:10.1136/jme.2003.007062

Sackett, D. (2003). Evidence-based medicine. What is it and what it isn't. *The Origins and Aspirations of ACP Journal Club*. Retrieved January 20, 2009, from http://www.minervation.com/cebm/ebmisisnt.html

Truman, B. I., Smith-Akin, C. K., & Hinman, A. R. (2000). Developing the Guide to Community Preventive Services--overview and rationale. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine*, *18*(1), 18–26. doi:10.1016/S0749-3797(99)00124-5