



## How to understand and conduct evidence-based medicine

Hyun Kang

*Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, Seoul, Korea*

Evidence-based medicine (EBM) is the conscientious, explicit, and judicious use of current best evidence in making decisions regarding the care of individual patients. This concept has gained popularity recently, and its applications have been steadily expanding. Nowadays, the term “evidence-based” is used in numerous situations and conditions, such as evidence-based medicine, evidence-based practice, evidence-based health care, evidence-based social work, evidence-based policy, and evidence-based education. However, many anesthesiologists and their colleagues have not previously been accustomed to utilizing EBM, and they have experienced difficulty in understanding and applying the techniques of EBM to their practice. In this article, the author discusses the brief history, definition, methods, and limitations of EBM. As EBM also involves making use of the best available information to answer questions in clinical practice, the author emphasizes the process of performing evidence-based medicine: generate the clinical question, find the best evidence, perform critical appraisal, apply the evidence, and then evaluate. Levels of evidence and strength of recommendation were also explained. The author expects that this article may be of assistance to readers in understanding, conducting, and evaluating EBM.

**Key Words:** Clinical epidemiology, Evidence-based medicine, Meta-analysis, Practice guideline, Systematic review.

Evidence-based medicine (EBM) refers to the conscientious, explicit, and judicious use of current best evidence in making appropriate clinical decisions [1]. Hence, it is necessary in EBM to connect the best external evidence with the values and preferences of patients and the expertise and insight of clinicians. Currently, EBM is attracting increasing interest in various fields, and consensus regarding the need for EBM is spreading, as EBM is considered to reflect the knowledge and skill that health care providers must possess. From a traditional point of view, medi-

cine is something that should be learned from a master, and the decisions made by clinicians regarding the diagnosis, treatment, prognosis, and risk factors of patients should be dependent only on clinical experience and practice. In addition, significant emphasis was placed on the pathophysiology of diseases, which alone was considered as a sufficient basis for decision-making, and the diagnosis and treatment based on an expert's opinion was considered to be a standard method of diagnosis and treatment. However, EBM encompasses the use of the results of systematic, reproducible, unbiased research as much as possible in clinical practice, as well as the implementation of patient-centered diagnosis and treatment on the basis of evidence, rather than diagnosis and treatment dependent solely on the judgment of medical doctors [2]. EBM seeks changes in the ways in which clinicians perform diagnosis and treatment, teach and learn medicine, and carry out research, which are summarized as follows: 1) Clinical practice should be conducted on the basis of best evidence, not relying on conventional methods; 2) Clinicians should treat patients with compassion, place patients at the center of diagnosis and treatment, and perform patient-oriented treatment; 3) Clinicians should learn or teach in a way

---

Corresponding author: Hyun Kang, M.D., Ph.D., M.P.H.  
Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, 102, Heukseok-ro, Dongjak-gu, Seoul 06973, Korea  
Tel: 82-2-6299-2586, Fax: 82-2-6299-2585  
E-mail: roman00@naver.com  
ORCID: <http://orcid.org/0000-0003-2844-5880>

Received: July 5, 2016.  
Accepted: July 15, 2016.

Korean J Anesthesiol 2016 October 69(5): 435-445  
<http://dx.doi.org/10.4097/kjae.2016.69.5.435>

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © the Korean Society of Anesthesiologists, 2016

Online access in <http://ekja.org>

that solves clinical problems based on clinical problem; and 4) Clinicians should employ a stricter approach toward research to increase the reliability of results.

Most of us, as anesthesiologists, perform anesthesia, pain treatment, and patient management based on years or decades of individual experience or the knowledge acquired from medical school, internship and residency, seminars and workshops, or recent medical research articles. The evidence on which we base our practice may not be something obtained externally, but something originating from our own experience or beliefs, something out of date, or something of which the quality has not been assessed through independent peer review. Hence, the quality of anesthesia, pain treatment, and patient management may be enhanced only after deliberation regarding how to obtain the best appropriate evidence; how to evaluate the quality of evidence; and how to connect the evidence with patients' values and preferences, or the clinical expertise and insight of individual clinicians, followed by application of the results. In addition, application of the principles and methodologies provided by EBM to routine practices may improve the treatment provided to patients.

The following article outlines EBM, the steps and methods of EBM practice, and the limitations of this approach.

## History of EBM

In 1990, Gordon Guyatt of McMaster University introduced a new concept of "Scientific Medicine." This concept was a novel method of medical education that was based on the clinically important skill of evaluation suggested by Sackett. Later, Guyatt [3] proposed EBM as an essential course in resident training programs.

Feinstein proposed the concept of "Clinical Epidemiology" as a novel direction in medical education, which is a combination of the statistical methods of epidemiology and clinical inference. He criticized public health studies as lacking rigor with respect to specified hypotheses, bias, poor data, and unsound inference of cause [4]. Clinical epidemiology was later defined again by Sackett as "the application, by a physician who provides direct patient care, of epidemiological and biometric methods to the study of diagnostic and therapeutic process in order to effect an improvement in health" [5].

In a series of articles in the *Canadian Medical Association Journal*, Sackett et al. introduced "Critical Appraisal" as a new method of reading medical literature. They demanded that the viewpoint of not only the readers of literature but also the users of the information be taken into account. Accordingly, they introduced the concept of "clinical practice guideline," which can be used by all clinicians to understand and apply a specific piece of literature, and published a series of relevant articles in the

*Journal of the American Medical Association* [6].

After this, the Cochrane Collaboration was formed; the institution's name is a tribute to Archie Cochrane and his efforts to eliminate bias in clinical research through randomized controlled trials. The concept that Cochrane suggested in his book entitled *Effectiveness and Efficiency: Random Reflections on Health Services* is sometimes considered as the starting point of EBM [7].

Later, EBM was adopted by the British National Health Service (NHS) as the main goals and methods for development of medical system. Recently, the application of EBM has been extended to not only medicine but also evidence-based health care, evidence-based nursing, evidence-based mental health, and evidence-based policy.

## Definition

EBM is the process of systematically finding, appraising, and using results obtained from well designed and conducted clinical research to optimize and apply clinical decisions [8]. Hence, the basic concept of EBM is to select and acquire medical information and apply the selected information to individual patients in order to decide to perform treatment if there exists evidence that treatment may be of benefit, or, alternatively, not to perform treatment if there is no evidence that treatment may be of benefit or if there is evidence that treatment may cause harm. The purposes of this method are to reduce bias that may be caused by relying on evidence obtained from a study that is not based on evidence from the best scientific literature for clinical decision-making or from a study that could have been conducted more systematically; to compare the benefits and risks in clinical decision-making; and to appraise decisions by taking patients' preferences into account. Here, "best evidence" is obtained from randomized controlled trials, if available, from observational research if there is no available randomized controlled trial, or from unsystematic clinical observation or pathological rationale if there is no available observational research.

In 1996, Sackett et al. [2] defined EBM as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of the individual patients." This definition of EBM means that, "Health care providers should be able to deal with changes in medicine which is rapidly developing with new knowledge and treatment methods. For this, health care providers should be conscientious, explicit and judicious in making decisions with regard to the new knowledge and treatment methods to apply to not an ideal patient but the individual patients to whom they actually provide health care."

However, this definition was criticized by many on the grounds that it is virtually impossible to put into actual practice and that the application of the definition may produce many

studies of low quality. Therefore, the definition of EBM was revised in 2000 to comprise “the integration of best research evidence with clinical expertise and patient value” [1].

“Current best evidence” refers to clinically appropriate studies, which means that current best evidence may be obtained from not only basic science research, but also clinical studies regarding the accuracy and precision of patient-centered diagnostic tests, the predictive power of prognostic markers, and the efficiency and safety of treatment. In addition, “clinical expertise” is the skillfulness obtained from clinical experience and represents the capability of clinicians to make decisions effectively and efficiently and to harmonize patients’ needs and preferences in patient treatment. The revised and improved definition “reflects a systematic approach to clinical problem solving” and emphasizes the importance of considering patient values.

The definitions described above indicate that all three elements of clinical problem solving, “patient,” “clinician,” and “information” should be taken into account. The element of “patient” includes a patient’s values, preferences, expectations, and financial status. The element of “clinician” includes a clinician’s education, experience, current expertise, continued learning, and attitude. The element of “information” includes the clinical appropriateness of the evidence, its research support, and how up-to-date it is. These elements are not independent but share one another’s domains or affect one another. The common part of these three elements is where the problems we suggest and the clinical questions about the problems are included and the point at which EBM begins.

## Steps in EBM

Various models of EBM have been created for various clinical settings, but the general steps of EBM are as follows [9].

- 1) Generate Clinical Question
- 2) Find Best Evidence
- 3) Critical Appraisal
  - Appraisal of evidence for validity and usefulness
- 4) Apply the Evidence
  - Application of results in clinical practice
- 5) Evaluate
  - Evaluation of information, skill, and EBM procedures

## Generate clinical question

Generating clinical questions, the first step of EBM, is the most difficult and the most important step. Turning a specific question raised into an answerable question is the essential technique in EBM [10]. Since there is plenty of information available around us, it is rare to fail to solve a certain problem because of lack of information. We fail to solve a certain prob-

lem mostly because we do not accurately recognize the necessity of the problem, we lack the time to search relevant information, or we fail to access existing information as a result of not being familiar with search techniques. In addition, one of the most common reasons why an answer to a question is not obtained is because the question itself is unanswerable. Therefore, it is essential to make a question as answerable as possible for problem solving.

Generally, a good question is 1) relevant and specific, 2) distinctly communicated, 3) clear in the objective and necessity of inquiry, and 4) one that will reduce the time required to obtain the answer.

To generate a good question for EBM, a clinician should be 1) specific in recognition of a problem and clarification of the clinical topic, 2) problem-oriented, 3) patient-focused, 4) able to determine answers through literature, and 5) able to consider various aspects such as patient selection and subjects to compare.

Considering these points, it is recommended to specify a clinical question for EBM as much as possible by applying the standard of PICO, where ‘P’ denotes ‘population’ or ‘patients’, ‘I’ ‘intervention’ or ‘exposure’, ‘C’ ‘comparison’, and ‘O’ ‘outcome’.

After generating a clinical question, the category of the question should be verified in order to understand the type of information required by the question. Verification of the category helps to decide on the appropriate strategy for a literature survey and enables clarification of the component that is not clearly known and determination of the research design to be considered for the category. Common categories of questions are diagnosis, prognosis, therapy, and risk (factors), and the type of research for answering the question is dependent on the category. If a clinical question belongs to the category of diagnosis, the optimal research design to answer the question is not a randomized controlled trial but a cross-sectional study or a case-control study. If a clinical question belongs to the category of prognosis, a cohort study is effective because a follow-up study of patients who have been evaluated from the early stage of a disease is necessary. If the category is therapy, a randomized controlled study or a systematic literature review of randomized controlled studies is usually needed. If a clinical question belongs to the category of risk (factors), a randomized controlled study, a cohort study, or a case-control study is necessary.

Depending on the type of information, study design is added to PICO to constitute PICO-SD, which clearly describes the distinct establishment of a study question.

Table 1 shows an example of PICO-SD specified for a question established by the Propofol Task Force Team of the Korean Society of Anesthesiologists according to the “Clinical Guidelines of Propofol Sedation for Non-Anesthesiologists.” The clinical question is, “Can combination therapy with propofol

**Table 1.** Examples of PICO-SD Selection

PICO	Description	Example
<b>P: Patients or Population</b>	Information on what subject group do I need?	Patients undergoing sedation therapy
<b>I: Intervention or Exposure</b>	Results of which intervention or exposure do I need?	Combination therapy with propofol and another sedative
<b>C: Comparison or Control</b>	What is the alternative to compare where an intervention is not performed or another intervention method is applied?	Propofol monotherapy
<b>O: Outcomes</b>	What is the effect or result of the intervention?	Risk of adverse effect
<b>SD: Study Design</b>	What study design will be included?	Randomized controlled trial

and another sedative make the risk of adverse effects lower than that of propofol monotherapy in patients undergoing sedation therapy?”

## Find the evidence

The conditions for the evidence acquired in EBM are that the evidence should be attainable, obtained externally from research or from an expert, up-to-date, timely, of high quality, applicable to individual patients, and appropriately the best [11].

### Identify resources

There are various resources available to use to find evidence. These resources include individual experience; intuition or rationale; peers' viewpoints; publications such as books, reports, and journals; electronic databases; and the assistance of expert librarians. It should be noted that individual experience, intuition, rationale, or peers' viewpoints may provide not information regarding the methods or practices that should be performed or applied but rather information regarding either current methods or practices or those performed or applied in the past. In addition, information derived from these sources may not be based on evidence or may not have been verified, or the information may be out-of-date, rather than the most up-to-date information. However, clinicians or their peers who are experts in their specialized fields may have information which may not be found in databases or books, or may be aware of some important journals that are not currently included in databases. Therefore, individual experience and peers' viewpoints are also very important sources to utilize in finding the evidence.

### Literature survey database

International literature survey databases include MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials (CENTRAL), LILACS, CINAHL, PsycINFO, Google, and Web of Science. Literature survey databases available in Korea include KoreaMed, KMBase, KISS, and RISS.

Beginning with the U.S. National Library of Medicine (NLM) in 1966, MEDLINE, which is the initial representative of medical information databases, includes about 4,000 international

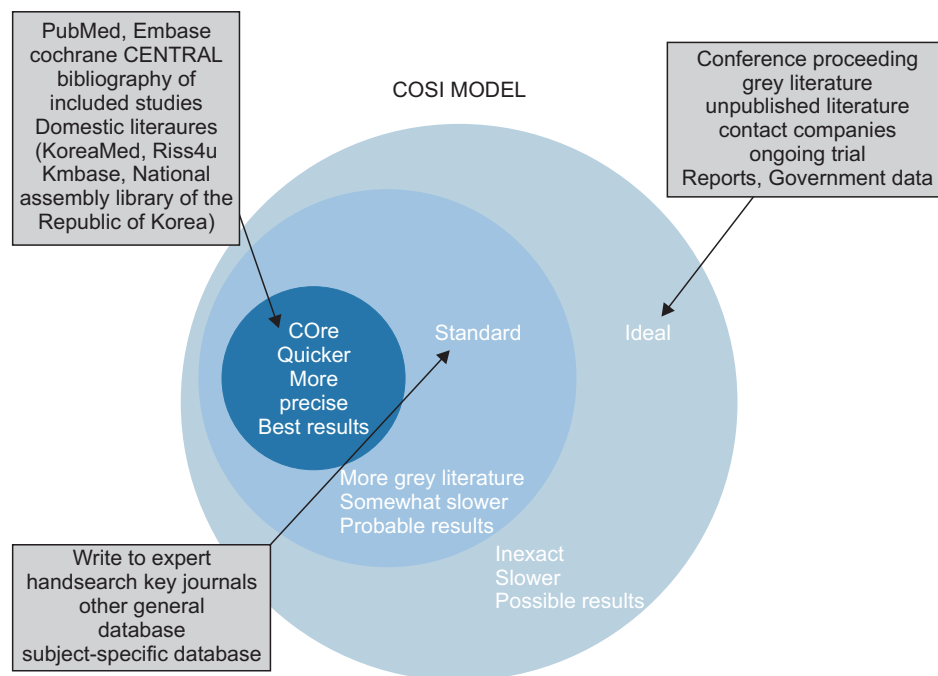
journals at present. MEDLINE includes the free PubMed database as well as the pay databases of Ovid (Ovid Technologies, Inc.) and WinSPIRS (SilverPlatter Information) provided by companies with special search strategies.

It is not that respective databases include different journals, but a certain journal may not be included in a certain type of database and may be included in multiple types of databases simultaneously. It is known that the scope of data and the scope of incorporation of MEDLINE and EMBASE are very different depending on the topic, but it has also been reported that about 30% of journals are redundant with respect to a certain topic [12]. Therefore, in order not to miss a piece of literature that should be found, a literature survey should be performed with the inclusion of all databases available. However, it may be virtually impossible to search all databases since too much effort and time are required. Therefore, the U.S. NLM provides the scope for a literature survey, which is the COSI (Core, Standard, Ideal) model [13] (Fig. 1). The term “core” means the essence of a literature survey, which is the minimal database required to find the best results rapidly and simply. Therefore, the “core” database is what must be searched. The group of “core” databases includes MEDLINE, EMBASE, and CENTRAL, while it also includes Korean databases such as KoreaMed, KMBase, KISS, and NDSL. The term “standard,” representing the standard scope of a literature survey, includes a manual search of core journals and the search of databases that are not “core” (Web of Science, DARE, CINAHL, and PsycINFO). The “ideal” part includes conference proceedings, gray literature, unpublished articles, and clinical trials currently in process [13].

Some think that search of the literature outside the scope of the “core” with the goal of increasing the sensitivity may reduce publication bias, but others think that the overall bias may be enhanced since the studies outside the scope of the “core” have not undergone peer review.

### Literature survey strategies

While there are countless pieces of information around us, the effort and time we may put into addressing any question are limited. Therefore, it is necessary to understand various survey strategies and establish an appropriate survey strategy not only



<http://www.nlm.nih.gov/archive/20060905/nichsr/ehta/chapter3.html#COSI> Fig. 1. COSI model.

to identify as many pieces of information as possible but also to rapidly detect useful information, and at the same time exclude baseless or inappropriate information.

Survey strategies are classified as strategies to increase sensitivity and strategies to increase specificity. Sensitivity is defined as the possibility of identifying relevant studies, searching for all the articles relevant to the topic without missing one. A survey with high sensitivity is a comprehensive survey that necessarily includes irrelevant articles. Specificity is defined as the possibility of excluding irrelevant studies, and a strategy to rule out articles that are not relevant to the topic. A strategy to increase sensitivity is aimed at incorporating important studies on a certain topic, and a survey focused on a research topic is one example. Harmonizing sensitivity and specificity is required in a survey strategy. However, a survey strategy to increase sensitivity is used for literature surveys for research, systematic literature reviews, and the development of clinical practice guidelines, whereas a survey strategy to increase specificity is chosen for literature surveys for knowledge or EBM [14].

A manual survey is also necessary in literature surveys in addition to searches using databases. In a manual survey, the references cited by the literature retrieved in the database survey may be reviewed, or the Science Citation Index may be used to search relevant literature. Since an important journal in the current field may not be included in the database to be surveyed, an expert in the field may be consulted. If an important journal is not included in the database, a manual survey should be per-

formed with that journal.

Additionally, surveys may be performed with gray literature, which means literature that has not undergone peer review. Gray literature may include reports and clinical trial registries. Databases that specialize in gray literature include Grey, NTS, and PsycEXTRA [13].

### Critical appraisal

Once the literature has been surveyed, appropriate literature should be selected and the quality and usefulness of the literature should be appraised. The quality of a study is appraised by considering the validity, reliability, and clinical importance of the study. The level of evidence of a piece of literature is determined by the quality appraisal, which in turn affects the strength of recommendation.

#### Selection of Literature

Since not all the searched literature is relevant to a particular study question, and many pieces of irrelevant literature are also retrieved by literature survey, the literature relevant to the study question is chosen by applying inclusion and exclusion criteria. Selection of literature goes through many steps. First, redundancy of the literature retrieved from various databases is checked. A bibliographical information management software program such as EndNote or Reference Manager is used to check the redundancy. However, since such bibliographical information



management software programs check the redundancy mechanically, pieces of literatures that are actually redundant may be considered different. Therefore, the redundancy should be verified manually. A pilot test may be performed before starting the selection and exclusion of the main body of literature selection to reduce the discrepancy between researchers regarding inclusion and exclusion as well as trial-and-error.

Objectivity and transparency should be maintained in the process of literature selection and exclusion so that the bias of individual appraisers may not intervene. The process of literature selection and exclusion should be carried out independently by at least two appraisers. In case of differing opinions on the part of appraisers, a conflict resolution strategy should be prepared in advance, including discussion between appraisers or intervention of a third party or an expert committee. In addition, reproducibility should be secured in the process of literature selection [15].

There is not a standardized format for the literature selection and exclusion process. However, generally, a primary selection procedure is carried out only with the titles and abstracts, followed by a second selection procedure to search the full text. When it is difficult to make decisions using only the titles and abstracts in the primary selection procedure, the full text should be searched to decide whether to include or exclude a piece of literature. If ambiguous, the literature should not be arbitrarily excluded by an appraiser. In the primary exclusion procedure, only the reason of exclusion and the number of pieces of excluded remain in the flowchart. A study that has not been excluded by at least one of the appraisers in the primary selection and exclusion procedure should undergo the secondary selection procedure. Selection is made in the secondary selection procedure on the basis of the full text of pieces of literature. During this procedure, reasons for excluding a certain piece of literature should be provided, and the list of excluded pieces of literature should be presented.

### Appraisal of quality of literature

After the selection of literature, the quality of the selected pieces of literature should be appraised. Appraisal of the quality of literature enables decisions to be made regarding whether to accept the conclusions drawn by the literature or not, conflicting results from different pieces of literatures to be interpreted and judgments to be made about them, and the need for additional studies to be determined. Although there are various methods of appraising the quality of literature according to study designs, not all of them are frequently used for various reasons, including convenience, validity, and difficulties in the application of relevant tools. Tools of appraisal are classified as scoring system and checklist system.

Generally, in the scoring system, scores for each appraisal

item are assigned and the total score is calculated. However, the meaning of the score is unclear and excessively simplified, the result is dependent on the type of tool, and the item regarding blinding randomized allocation is not considered in the scoring system. In the scoring system, the Chalmers scale and the Jadad scale are used. In the checklist system, the checklist developed by the Scottish Intercollegiate Guidelines Network (SIGN), the Risk of Bias Assessment tool for Nonrandomized Studies (RoBANS), and the Risk of Bias tool developed by Cochrane's group are used.

The methods of appraising the quality of randomized controlled studies include the Chalmers scale, the Jadad scale, the checklist developed by SIGN, and the Risk of Bias tool. The methods of appraising the quality of nonrandomized controlled studies include the Newcastle-Ottawa scale, the Deeks criteria, MINORS, and RoBANS.

Table 2 shows the items included in the Risk of Bias tool developed by Cochrane's group to appraise the quality of randomized controlled studies. These items are known to have a significant effect on the results of a study. While the tool has simplified the items, it has also reduced the possibility of subjective and arbitrary answers to the same items as well as the possibility of lowered reliability depending on the appraiser's understanding and skillfulness with respect to the methodology. In addition, the Risk of Bias tool is easy to use because it provides specific de-

**Table 2.** Risk of Bias Table

Item	Author's judgment
<b>Random sequence generation</b>	<input type="checkbox"/> Yes
Selection bias (biased allocation to interventions) due to inadequate generation of a randomized sequence.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Allocation concealment</b>	<input type="checkbox"/> Yes
Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Blinding of participants and personnel</b>	<input type="checkbox"/> Yes
Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Blinding of outcome assessment</b>	<input type="checkbox"/> Yes
Detection bias due to knowledge of the allocated interventions by outcome assessors.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Incomplete outcome data</b>	<input type="checkbox"/> Yes
Attrition bias due to amount, nature, or handling of incomplete outcome data.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Selective reporting</b>	<input type="checkbox"/> Yes
Reporting bias due to selective outcome reporting.	<input type="checkbox"/> No <input type="checkbox"/> Unclear
<b>Other bias</b>	<input type="checkbox"/> Yes
Bias due to problems not covered elsewhere in the table.	<input type="checkbox"/> No <input type="checkbox"/> Unclear

scriptions and guidelines for each item to select from one among “high risk of bias, No,” “low risk of bias, Yes,” and “uncertain risk of bias, unclear” depending on the information included in the literature [16].

### Bias

The quality of literature is represented by the risk of bias. All measurement values different from the true values are called error, and errors are classified as random error, which is inconsistent, and systematic error, which has some direction. Random error is generated regardless of the research procedure or method, but systematic error, also called bias, is generated in cases in which there is a problem in the research procedure or method. There are various types of bias (Table 3). A biased study may provide an erroneous conclusion or may overestimate or underestimate results. Therefore, a study methodology should be strictly applied to minimize the risk of bias [17]. The presence of bias itself may not be measured, and only the risk of bias may be evaluated. Therefore, the quality of literature may be appraised by evaluating the risk of bias. Risk of Bias is also the name of a tool for appraising the quality of randomized controlled studies.

Methods of interpreting actual analytical results in the presence of risk of bias include the method of analyzing all the studies and describing the risk of bias, the method of including only the studies having a low risk of bias in the analysis, and the method of presenting various analytical results including the analytical results from all studies as well as the analytical results only from the studies having a low risk of bias in the analysis.

### Summary of evidence

Data analysis is the step in which extracted data are analyzed, summarized, and synthesized. Data synthesis is divided into qualitative synthesis and quantitative synthesis. When statistical quantitative synthesis is impossible, only qualitative synthesis is performed to describe and present respective results. When statistical synthesis is possible, quantitative synthesis is performed.

**Systematic Review:** Systematic literature review refers to collecting all available studies by using an objective, systematic, and reproducible methodology to answer a specific and clearly de-

scribed study question; searching and selecting studies by using a clear and systematic method to obtain valid and reliable results with minimal bias; and presenting the results obtained from the selected studies after appraising the validity of the results and discussing and analyzing the results.

The characteristics of systematic literature review, in comparison with narrative review, are 1) that the study selection criteria are provided in advance and the objectives are clearly established, 2) that a clear and reproducible methodology is used, 3) that a systematic attempt is made to identify all studies satisfying the selection criteria, 4) that the validity of the selected studies is measured, and 5) that the results and characteristics of the selected studies are systematically presented and synthesized [18].

**Meta-analysis:** Meta-analysis refers to a statistical method of synthesizing pooled estimates by summarizing estimates from two or more individual studies. In other words, meta-analysis is a statistical method of quantitatively calculating pooled estimates by integrating results from studies and evaluating the effect and efficiency of the pooled estimates [19]. Therefore, meta-analysis is also called analysis of analyses. In general, systematic literature review and meta-analysis are performed simultaneously. However, it is not necessary to perform a meta-analysis when a systematic literature review is performed: a meta-analysis may or may not be performed. For example, if the characteristics of studies are too heterogeneous, quantitative synthesis may not be attempted. In contrast, a meta-analysis, although it is generally performed following a systematic literature review, may be performed without performing a systematic literature review.

The purpose of a meta-analysis is not just to find out summary estimates. If there is a consistent pattern among research results, it is important to determine the meaning. If there is not a consistent pattern, it is important to ascertain what makes the results inconsistent. The characteristic of being inconsistent is called heterogeneity. In meta-analysis, heterogeneity refers to cases in which the variation observed in the results integrated from individual studies by a meta-analysis is greater than the variation of sampling, so the variation may not be attributed to chance. Heterogeneity may be caused by the variety of clinical settings among studies, the variation in methodologies, chance, and bias.

The presence and degree of heterogeneity may be verified by a visual test using a plot or by a statistical test. The most representative visual test method using a plot to verify heterogeneity is to draw a forest plot and verify whether there is an overlap between the directionality of the therapeutic effect values of individual studies and the confidence interval. Other methods of using a plot employ the L'Abbe plot and the Galbraith plot. Statistical tests for verifying heterogeneity include the  $\chi^2$  test (Q statistics) [20] and Higgins'  $I^2$  statistic [21].

**Table 3.** Kinds of Biases

<b>Selection bias</b>
Medical surveillance bias, non-response bias, Berkson's bias, selective survival bias, volunteer bias (self-selection bias), follow-up loss bias
<b>Information bias</b>
Interviewer bias, performance bias, measurement bias, memory decay bias, publication bias, Hawthorne bias, ascertainment bias, attrition bias, time bias
<b>Confounding bias</b>

**Bias in Meta-Analysis:** Bias in meta-analysis includes publication bias, language bias, location bias, and time lag bias. The most representative bias is publication bias, which refers to the bias in meta-analysis caused by the fact that studies showing statistically significant results are included more often in meta-analysis because those studies are more likely to be published than studies that do not show statistically significant results. Bias may cause overestimation or underestimation of study results or lead to an erroneous conclusion, resulting in a wrong interpretation or difficulties in interpretation.

The presence and degree of publication bias may be verified by a visual test using a plot or by a statistical test. The most representative method of using a plot is to employ a funnel plot. In a funnel plot, the y-axis represents the sample size of a study and the x-axis represents the effect size. Individual studies are expressed as points in the scatter diagram having the shape of an upside-down funnel, since the top is narrow while the bottom is wide. The methods based on statistical tests include Begg and Mazumdar's rank correlation test [22] and Egger's test [23].

### Levels of evidence

Evidence refers to what is proved by studies conducted according to the best research methodology, and level of evidence refers to the degree of confidence in the effect of intervention on the basis of the current evidence. Level of evidence is expressed in various other terms such as quality of evidence by the Grading of Recommendations Assessment, Development and Evaluation (GRADE), level of uncertainty by the U.S. Preventive Services Task Force (USPSTF), strength of evidence by the Agency for Healthcare Research and Quality (AHRQ), and quality of a body of evidence by Cochrane.

Studies are conducted with various types of designs, and the level of clinical evidence is determined by the risk of various biases that may be present in a particular type of design. For example, it is considered that a randomized controlled, double-blind, placebo-controlled trial conducted with a homogeneous patient group and completely followed up provides the least risk of bias and the strongest evidence. On the contrary, a case report or an expert opinion is considered to have a weak level of evi-

dence because it has a high probability of bias.

Many other factors contribute to the determination of the level of evidence, including quality of literature, quantity of evidence, consistency of evidence, and directness of evidence. Table 4 displays the level of evidence system provided by SIGN as one of the scales. This system provides levels of evidence on the basis of study design and quality of literature. The studies located near the top of the table are expected to have a low risk of bias and a high level of evidence. This system, widely used in the past, is not used currently because it is known that the system of grading the recommendations in A, B, C, and D according to the level of evidence is not appropriate. Therefore, SIGN decided to accept the internationally recognized GRADE system in 2013.

The strength of evidence may be determined by the quantity of evidence, consistency of evidence, and directness of evidence. Quantity of evidence is determined by the number of research articles on the topic, the number of subjects, and the effect size. Consistency of evidence is determined by how similar the results of individual studies are in terms of the direction and size of effect.

### Strength of recommendation

The strength of recommendation of an intervention refers to the level of confidence in obtaining a desired result by performing the intervention according to the recommendation. The strength of recommendation is determined by considering the level of evidence and the balance between the benefit and risk of the intervention. The GRADE working group prepared a system of strength of recommendation by considering not only the quality of medical studies but also the viewpoint affecting the reliability of results. In this system, the quality of evidence is appraised by grading into four categories of "high," "moderate," "low," and "very low," according to how an effect observed across the entire body of evidence, not just in individual research results, may be similar to the actual effect. The "high" grade means that further research is very unlikely to change the confidence in the estimate of an effect. The "moderate" grade means that further research is likely to have an important impact on the confidence in the estimate of an effect and may change the esti-

**Table 4.** SIGN Level of Evidence (1999–2012)

1++	High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well-conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1–	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High-quality systematic reviews of case-control or cohort or studies
2+	High-quality case-control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2–	Well-conducted case-control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
3	Case-control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, e.g., case reports, case series
4	Expert opinion

RCTs: randomized controlled trials.



mate. The “low” grade means that further research is very likely to have an important impact on the confidence in the estimate of an effect and is likely to change the estimate. The “very low” grade means that any estimate of effect is very uncertain.

Initially, the “high” grade is assigned to a randomized study, and the “low” grade is assigned to an observational study. The grade is lowered if there is a risk of bias in the research design, if the evidence lacks directness, if the accuracy of the research is low, if there is unexplained heterogeneity, or if there is a risk of publication bias. The level of recommendation is elevated if the effect size is great, if there is a low possibility of confounders, or if there is a dose-response. The grade of a randomized controlled study is not elevated [24].

### Apply the evidence

Following literature survey and evidence appraisal, the results determined by evaluation to be useful should be applied in actual clinical settings. A complicated, well-conducted study presented in a sophisticated article will not be helpful if the results are not applied to patients. In addition, if new discoveries from studies are applied in clinical settings too slowly or not applied at all, the potential benefit that clinical studies may provide to clinical settings will be removed. However, it seems that the application of evidence is not receiving sufficient attention. In fact, there are many studies in which data analysis was conducted to improve the quality of medical service and for patient safety, but no studies exist regarding how to apply evidence. The application of evidence itself is very difficult and requires changes in the medical system, individual clinicians, and ultimately, the entire culture of the medical system.

Evidence obtained from EBM may be applied to various situations, but that application neither always occurs in medical services nor results in change [25]. Even when EBM is applied, the application is often inappropriate. According to one report, even in the situations in which EBM was applied for patient treatment, 18% of the doctors changed the patient therapy when recommendations were presented as the results of a literature survey [26].

If the research has been conducted adequately, EBM may be performed. However, there are cases in which sufficient research has not been conducted. In such cases, decisions are made on the basis of evidence that is not based on research (e.g., expert opinion or scientific inference) [27]. In these cases, the clinician, as a researcher, may need to make efforts to produce evidence.

Since no patients are identical, they have different values, preferences, expectations, and circumstances. Patients are often encountered in a situation that is different from the one that has been searched and evaluated. Therefore, applying the “ideal” evidence that was retrieved and evaluated may not be the best for

the patient to be treated. In this case, decisions should be made by considering the amassed evidence as well as the situation of the patient.

In addition, clinicians having different types of training, experience, and specialties may prefer different treatment methods. Therefore, the retrieved evidence may conflict with the treatment method prioritized by each clinician. Conflicts may occur in such cases. In these cases, EBM may help the patient make a decision about intervention or treatment.

### Evaluate

After the application of evidence, the information, intervention, and EBM process are evaluated. The evaluation is performed with respect to the appropriateness of the quantity and quality of evidence in the information, the difficulty of acquiring the evidence and results, the cost of application, the response and the degree of compliance on the part of patients, the difficulty of application, the clinical results, the effect of the actual application, and the changes that the experience has produced in the thoughts and skills of clinicians.

A feedback mechanism should be in place for the knowledge acquired in the process of actually applying the evidence, so that others may perform the process well and the EBM implementation strategy may be improved.

### Strategy for Rapid Decision-Making

The application of EBM may be difficult if a clinical decision should be made rapidly. Sacket and Strauss [28] tested whether it is possible for clinicians, during their rounds, to use the “evidence cart” containing various types of evidence to search for and apply relevant evidence. This study was conducted due to the concern that actual EBM application may be limited because time and effort are required for a systematic literature appraisal before decision-making. The study showed that the “evidence cart” was often applicable as rapidly as medical service may be provided during team rounds, and that the “evidence cart” could affect decisions about diagnosis and treatment in 81% of cases, among which 91% had successful outcomes for patient treatment. Therefore, it was stated that the “evidence cart” is helpful for busy clinicians to perform rapid and appropriate decision-making. Rapid decision-making may be helped if there is a protocol such as practice guidelines for quick reference or if the EBM library is easily accessible. Therefore, efforts are continuously made to prepare and implement practice guidelines, while the portal sites of treatment guidelines also provide practice guidelines (e.g., NGC [National Guideline Clearinghouse, <http://www.guideline.gov>], GIN [Guidelines International Network, <http://www.g-i-n.net>]).

An EBM library enabling quick access and search should be convenient to use, accept up-to-date information, and include electronic databases. The EBM libraries satisfying these conditions include medical information services such as Medscape and HDCN and review services such as Evidence-Based Medicine Review, Cochrane Library, Best Evidence, and Up to Date.

## Limitations

The limitations of EBM are that 1) logical and consistent scientific evidence may be insufficient, 2) applying the retrieved evidence to the treatment of a specific patient is often difficult, 3) there is a barrier in applying high-quality medical skills, 4) clinicians often do not have the skills required for literature survey or appraisal, 5) busy clinicians have limited time to master and apply EBM, and 6) the resources needed to find the evidence are often inadequate in clinical settings [2].

This article has explained the general outline and history of EBM, the steps and methodologies of EBM, the strategy for rapid decision-making, and the limitations of EBM.

The term “EBM” is broadly used in the preparation of practice guidelines and the education and policy-making of EBM implementation. Recently, the term has been widely used in the domains where evidence is emphasized on the level of the group

or the individual. In addition, EBM is continuously undergoing advancements and revisions. Although EBM has been developed in many areas and applied to actual clinical settings, it has not been perfectly applied to medicine, a field in which experience is highly regarded [29]. Moreover, many clinicians feel it is difficult to understand and perform EBM.

EBM is like a double-edged sword to clinicians. While EBM may ensure that the diagnosis and therapy provided is based on scientific evidence, it reduces the space in which individual clinicians may perform diagnosis and therapy on the basis of their clinical judgment. Just as a person who has mastered how to use a knife in a wrong way may become a robber but a person who has mastered its use in a right way may be an excellent medical doctor, a clinician who masters EBM rightly and introduces it will become an excellent clinician, while the clinician may do harm to himself or herself or to patients if he or she does not understand and use EBM appropriately. In addition, since the global trend is toward the emphasis and application of EBM, clinicians who do not apply it will confront challenges from others. Therefore, just as an excellent doctor masters the use of a surgeon's knife, we need to master the interpretation and application of EBM to harmonize our knowledge and capability with the global advances in knowledge.

## References

1. Sackett DL. Evidence-based medicine: how to practice and teach EBM. London, Churchill-Livingstone. 2000, p 1-20.
2. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-2.
3. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 114: A16.
4. Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. *Science* 1988; 242: 1257-63.
5. Sackett DL. Clinical epidemiology. *Am J Epidemiol* 1969; 89: 125-8.
6. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268: 2420-5.
7. Cochrane A. Effectiveness and Efficiency: Random Reflections on Health Services. London, Royal Society of Medicine Press. 1999, pp 1-92.
8. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* 1995; 310: 1122-6.
9. Titler MG. Developing an evidence-based practice. In: *Nursing Research: Methods and Critical Appraisal for Evidence-based Practice*. 7th ed. Edited by LoBiondo-Wood G, Haber J: St. Louis, MO: Mosby. 2006, pp 385-437.
10. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12-3.
11. Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *J R Coll Physicians Lond* 1998; 32: 557-63.
12. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Control Clin Trials* 2000; 21: 476-87.
13. Bidwell S, Jensen MF. Chapter 3: Using a Search Protocol to Identify Sources of Information: the COSI Model. In: *Etext on Health Technology Assessment (HTA) Information Resources* [Internet]. Bethesda: U.S. National Library of Medicine, National Institutes of Health, Health & Human Services; 2003 Jun 14 [cited 2016 Jul 5]. Available from <http://www.nlm.nih.gov/archive/20060905/nichsr/eha/chapter3.html#COSI>.
14. Marlborough HS. Accessing the literature: using bibliographic databases to find journal articles. Part 1. *Prim Dent Care* 2001; 8: 117-21.
15. Higgins JP, Deeks J. Chapter 7: Selecting studies and collecting data. In: *Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration*. 2011 [updated 2011 Mar; cited 2016 Jul 5]. Available from <http://handbook.cochrane.org>.

16. Higgins JP, Altman DG, Sterne JA. Chapter 8: Assessing the risk of bias in included studies. In: *Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration*. 2011 [updated 2011 Mar; cited 2016 Jul 5]. Available from <http://handbook.cochrane.org>.
17. Lee S, Kang H. Statistical and methodological considerations for reporting RCTs in medical literature. *Korean J Anesthesiol* 2015; 68: 106-15.
18. Green S, Higgins JP, Alderson P, Clarke M, Mulrow CD, Oxman AD. Chapter 1: Introduction. In: *Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration*. 2011 [updated 2011 Mar; cited 2016 Jul 5]. Available from <http://handbook.cochrane.org>.
19. Kang H. Statistical considerations in meta-analysis. *Hanyang Med Rev* 2015; 35: 23-32.
20. Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials* 1986; 7: 267-75.
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-60.
22. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088-101.
23. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-34.
24. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64: 401-6.
25. Ward MM, Evans TC, Spies AJ, Roberts LL, Wakefield DS. National Quality Forum 30 safe practices: priority and progress in Iowa hospitals. *Am J Med Qual* 2006; 21: 101-8.
26. Lucas BP, Evans AT, Reilly BM, Khodakov YV, Perumal K, Rohr LG, et al. The impact of evidence on physicians' inpatient treatment decisions. *J Gen Intern Med* 2004; 19: 402-9.
27. Titler MG, Kleiber C, Steelman VJ, Rakel BA, Budreau G, Everett LQ, et al. The Iowa model of evidence-based practice to promote quality care. *Crit Care Nurs Clin North Am* 2001; 13: 497-509.
28. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". *JAMA* 1998; 280: 1336-8.
29. Eddy DM, Billings J. The quality of medical evidence: implications for quality of care. *Health Aff (Millwood)* 1988; 7: 19-32.