2
Evidence-Based Surgery Defined

“It is the mark of an educated mind to be able to entertain a thought without accepting it.”
— Aristotle

Summary

In this chapter, evidence-based medicine (EBM) is defined and the necessity and challenges of practicing EBM in surgery are presented. Resources are provided to aid the surgeon in obtaining the best available evidence.

Introduction

The term “evidence-based medicine” (EBM) first appeared in autumn 1990 in a document for applicants to the Internal Medicine Residency Program at McMaster University that described EBM as an attitude of “enlightened skepticism” toward the application of diagnostic, therapeutic, and prognostic technologies. As outlined in the text Clinical Epidemiology¹ and first described in the literature in the ACP Journal Club in 1991,² the EBM approach to practicing medicine relies on an awareness of the evidence upon which a clinician’s practice is based and the strength of inference permitted by that evidence. The most sophisticated practice of EBM requires, in turn, a clear delineation of relevant clinical questions, a thorough search of the literature relating to the questions, a critical appraisal of available evidence and its applicability to the clinical situation, and a balanced application of the conclusions to the clinical problem. The EBM model integrates research evidence, clinical circumstances, patients’ values/preferences, and clinical experience (Fig. 2.1).

Fig. 2.1 Current model of evidence-based medicine.

How Evidence-Based Medicine Differs from Traditional Approaches to Health Care

According to the traditional paradigm, clinicians evaluate and solve clinical problems by reflecting on their own clinical experience or the underlying biology and pathophysiology or by consulting a textbook or local expert. For many traditional practitioners, reading the Introduction and Discussion sections of a research article is sufficient for gaining relevant information, and observations from day-to-day clinical experience are a valid means of building and maintaining knowledge about patient prognosis, the value of diagnostic tests, and the efficacy of treatment. Because this paradigm places high value on traditional scientific authority and adherence to standard approaches,³ traditional medical training and common sense provide an adequate base for evaluating new tests and treatments, and content expertise and clinical experience are sufficient to generate guidelines for clinical practice.

Evidence-based practice posits that although pathophysiology and clinical experience are necessary, they alone are insufficient guides for practice. These evidence sources may lead to inaccurate predictions about the performance of diagnostic tests and the efficacy of treatments. Like the traditional approach to health-care, the evidence-based...
health-care paradigm also assumes that clinical experience and the development of clinical instincts (particularly with respect to diagnosis) are crucial elements of physician competence. However, the EBM approach includes several additional steps. These steps include using experience to identify important knowledge gaps and information needs, formulating answerable questions, identifying potentially relevant research, assessing the validity of evidence and results, developing clinical policies that align research evidence and clinical circumstances, and applying research evidence to individual patients with their specific experiences, expectations, and values.\(^5\)

### Key Concepts: The Five As of Evidence-Based Medicine

1. **Ask** – Formulate your question.
2. **Acquire** – Conduct an efficient search for the best available research evidence.
3. **Appraise** – Is the evidence you found valid?
4. **Apply** – Use the best available evidence and decide whether it is applicable to your specific patient question.
5. **Act** – When evidence is valid, take what you have learned back to your patient.

Unfortunately, practicing EBM is not easy. Practitioners must know how to frame a clinical question to facilitate use of the literature in its resolution. Typically, a question should include the population, the intervention, and relevant outcome measures. The question, “What is the role of internal fixation of tibial fractures?” is vague. The question should be “In patients presenting to the emergency room with open tibial diaphyseal fractures (population), what is the effect of external fixators versus nonreamed intramedullary nails (interventions) on reoperation rates (outcome)?”\(^6\)

EBM practitioners (i.e., clinicians who work under the EBM paradigm) regularly consult original literature, including the Methods and Results sections of research articles.\(^6\) Correctly interpreting literature on prognosis, diagnostic tests, and treatment and potentially harmful exposures (medications’ side effects, environmental exposures) requires an understanding of the hierarchy of evidence. For example, in making treatment decisions, EBM practitioners may conduct an n-of-1 randomized trial (randomized trial in an individual patient, with the patient repeatedly treated with active intervention or placebo) to determine the optimal treatment for an individual patient.\(^7\) Alternatively, they may seek a systematic review of randomized trials of treatment alternatives. If a systematic review is not available, they will look for individual randomized trials and high-quality observational studies of relevant management strategies. If the literature is lacking altogether, EBM practitioners will fall back on the underlying biology and pathophysiology, and clinical experience.

### The Need for Evidence-Based Medicine

Over the last several years, the concepts and ideas attributed to and labeled collectively as evidence-based medicine have become a part of daily clinical lives, and clinicians increasingly hear about evidence-based guidelines, evidence-based care paths, and evidence-based questions and solutions. The controversy has shifted from whether to implement the new concepts to how to do so sensibly and efficiently, while avoiding potential problems associated with several misconceptions about what EBM is and what it is not. The EBM-related concepts of hierarchy of evidence, meta-analyses, confidence intervals, study design, and so on, are so widespread, that clinicians to understand today’s medical literature have no choice but to become familiar with EBM principles and methodologies.

The skills associated with EBM should allow clinicians to function more rationally. The ability to follow the path from research to application should also provide more control over what we do, and more satisfaction from our daily practice. Although learning to locate, assess, and use new evidence in the original literature can improve our daily practice, limited access to that information and limited time allocated to continuing education may cause our up-to-date clinical knowledge to deteriorate with time. EBM-related skills provide solutions to deal with this problem by allowing us to access, appraise, and apply information much more efficiently.\(^9\)

Critics of EBM have mistakenly suggested that EBM equates evidence with results of randomized trials, statistical significance with clinical relevance, evidence (of whatever kind) with decisions, and lack of evidence of efficacy with the evidence for the lack of efficacy. Other critics argue that EBM is not a tool for providing optimal patient care, but merely a cost-containment tool.\(^10\) All these statements represent a fundamental mischaracterization of EBM.
managers, nurses, and clerical staff with the purpose of obtaining costs. Detailed cost data collection sheets were prepared to calculate costs, which included staff costs, consumables including drugs, intravenous fluids, equipment, investigations, laundry, catering, and stationery. An inventory of costs per item was obtained, and the total cost was calculated from the number of items used. Outcomes were measured in terms of survival. The total costs of emergency AAA repair were pounds sterling 96,700.69, with a cost per life saved of pounds sterling 24,175.17. The total cost of elective AAA repair was pounds sterling 76,583.22, with a cost per life saved of pounds sterling 5,470.23. Emergency intervention for AAA was found to cost five times more than a planned intervention per life saved per year.

Cost-Utility Analysis

In a cost-utility analysis (CUA), the effectiveness of a surgical procedure can be measured in terms of lives saved, limbs saved, or days off work averted. Such measurement outcomes, however, do not allow one to easily compare the benefits across different types of medical interventions, for example, coronary bypass versus limb transplantation. Cost-utility analysis is able to incorporate the increase in the health-related quality of life (Fig. 13.5). Third-party payers and policy makers would prefer this type of presentation of effectiveness as they need to decide where to allocate scarce health care resources.

A cost utility analysis is a type of cost-effectiveness analysis in which the consequences are expressed in terms of life-years adjusted by peoples’ preferences; typically, one considers the incremental cost per incremental gain in quality-adjusted life-years.

Jargon Simplified: Cost-Utility Analysis

"A cost-utility analysis is a type of cost-effectiveness analysis in which the consequences are expressed in terms of life-years adjusted by peoples’ preferences; typically, one considers the incremental cost per incremental gain in quality-adjusted life-years."

Cost-utility analysis addresses these limitations by using a common outcome measure (a metric) to account for the broad range of relevant outcomes. In particular, the consequences of the intervention are expressed as utilities, which can be viewed as the preferences of individuals or society may have for a particular health state referred to as utilities. Utility is usually expressed as a decimal from zero to one, with zero representing death and one representing perfect health. There are different approaches to utility measurement. Utilities (preferences) are global health-related quality of life (HRQL) measures. They can be obtained from a visual analog scale such as the “feeling thermometer,” standard gamble, time-tradeoff, or from generic scales such as the Health Utilities Index (HUI) or the EuroQol-5D (EQ-5D). These methods are described in detail by Drummond et al.

One of the most commonly used utility measures in CUA is quality-adjusted life-years (QALYs) with results of the analysis expressed as cost per QALY gained. Briefly, QALYs are calculated by multiplying the life years gained from an intervention by the utility weight that can be determined by various methods. This common outcome measure allows for incorporation of both changes in quantity of life and quality of life.

Key Concepts: Incremental Cost-Utility Ratio (ICUR)

ICUR = $\frac{\Delta \text{Cost}}{\Delta \text{QALYs}}$

When determining whether to accept a new intervention following a CUA, the same principles apply as in the CEA (Fig. 13.3, Fig. 13.4). When a new intervention is both more effective and most costly guidelines exist to recommend whether to adopt or reject the new intervention. Interventions that cost less than $20,000 / QALY
Examples from the Literature: An Example of a Cost-Utility Analysis


**Abstract:** This study was a pragmatic economic evaluation performed alongside a multicentre randomized controlled trial comparing laparoscopic with open groin hernia repair. The primary economic evaluation framework employed was a cost-utility analysis. At 26 hospitals in the UK and Ireland, 928 patients with a groin hernia were assigned randomly to laparoscopic or open repair. Cost data were identified and measured both within and without the trial. Cost data were combined with quality-adjusted life-years (QALYs) from the EQ-5D questionnaire to obtain cost-per-QALY ratios. The mean cost of laparoscopic hernia repair was £1112.64, compared with £788.79 for the open operation. The extra cost of £323.85 in the laparoscopic group was mainly due to additional theater time and increased equipment and sterilization costs. The estimated incremental cost per QALY of the laparoscopic over the open method was £548.00 (95 per cent confidence interval £216.00–£885.00). While the results show that a high cost was incurred to produce an additional QALY by using laparoscopic over open hernia repair, sensitivity analyses show that there are specific situations in which laparoscopic repair may be a viable alternative, such as when reusable equipment is employed.

Cost-Benefit Analysis

The cost-benefit analysis (CBA) is a form of economic analysis in which the costs and the consequences (including increases in the length and quality of life) are expressed in monetary terms. This means that all benefits and costs of each intervention are measured in terms of their equivalent money value. A technique referred to as willingness to pay is used to assign the monetary value to the consequence of each outcome.

An advantage of the CBA is that it permits a direct comparison of various programs, as both costs and consequences are reported in the same units (dollars). The main criticism of the CBA in health interventions is that it may show bias toward the rich, if their willingness to pay were higher than that of the poor. For additional information on CBA, see Drummond et al.

**Jargon Simplified: Cost-Benefit Analysis**

“Cost-benefit analysis is a form of economic analysis in which the costs and the consequences (including increases in the length and quality of life) are expressed in monetary terms.”

Examples from the Literature: Example of a Cost-Benefit Analysis


**Abstract:** From a public health perspective, a cost-benefit analysis of using bike/pedestrian trails in Lincoln, Nebraska, to reduce health care costs associated with inactivity was conducted. Data was obtained from the city’s 1998 Recreational Trails Census Report and the literature. Per capita annual cost of using the trails was $209.28. Per capita annual direct medical benefit of using the trails was $564.41. The cost-benefit ratio was 2.94, which means that every 1 U.S. dollar investment in trails for physical activity led to 2.94 U.S. dollars in direct medical benefit. The sensitivity analyses indicated the ratios ranged from 1.65 to 13.40. Therefore, building trails is cost beneficial from a public health perspective. The most sensitive parameter affecting the cost-benefit ratios were equipment and travel costs; however, even for the highest cost, every 1 U.S. dollar investment in trails resulted in a greater return in direct medical benefit.

**Conducting an Economic Analysis**

The ideal economic analysis would consist of a large multicenter randomized controlled trial, which recruits a few thousand patients (based on a formal sample size calculation) comparing two interventions. Costs would then be collected prospectively along with the sampled data. When designing and conducting an economic analysis, one should consider collaborating with a health economist to ensure that the correct methodology is followed. Below are several additional items that need to be considered when conducting an economic analysis.
Classifying Types of Data

Data are the basic ingredients of a statistical analysis. One must understand their form to appropriately describe them and any relationships that lie therein. Data describe various characteristics of an observation (a patient, subject, case, or incident) and come in two main forms: categorical or numerical. The basic components are often used to calculate the frequency of occurrence of a particular event or outcome in the form of a rate or proportion. These types of data are described below.

Categorical Data

Also known as qualitative data, these data are described as having two or more categories. The simplest form of categorical data has two categories only and typically indicates the presence or absence of some attribute. Examples include:

1. Male/female
2. Dead/alive
3. Diabetic/nondiabetic
4. Fracture union/fracture nonunion
5. Implant survived/implant failed
6. Smoker/nonsmoker

These data are often called binary. Note that although gender and mortality are unequivocal categories, being classified as a smoker, diabetic, or having united a fracture relies on some more subjective cutoff point (number of cigarettes per day or number of cortices bridged), which requires a simplification of more complex data. Although the latter process is less favorable statistically because it involves a loss of information, it is often necessary for ease of interpretation. For example, it may be more desirable clinically to report the impact of a particular type of surgery on fracture union rates than number of radiographically bridged cortices even though the latter may be more quantitatively accurate.

When categorical data have more than two classes, they can either have some sort of natural ordering or not. Blood type (A/B/AB/O) or injury type (blunt/penetrating/burn) are two examples of data that lack a natural ordering or progression from one to the next category and are called nominal data. Various staging classification such as for cancer or grade of open fracture have an obvious ordering of categories and are referred to as ordinal data.

Numerical Data

Numerical data differ from categorical data in that they are quantitative. They come in two basic forms, discrete and continuous. Discrete data come about when a variable of interest can only take on a certain numerical value. It may be a certain number of visits to one doctor or the number of times a patient asks for pain medication. These are typically counts of events and differ from ordinal categorical data in that there is some additivity and proportionality to discrete data. Four visits to one's doctor are twice as many or two more than two visits to the doctor. Stage IV cancer is not necessarily twice as "bad" as stage II cancer, and the clinical difference between stage II and III is not necessarily the same as going from stage III to IV.

Continuous data are usually obtained by some form of measurement and are not limited to any value other than by the precision of the instrument. Height, weight, and blood pressure are all continuous data that can be reported to as many decimal places or fractions of a millimeter of mercury as are provided by the measurement device. Even though continuous data are often lumped into categories (i.e., ages 1 to 5, 6 to 15, 16 to 25, and so on), the richness of information is usually optimized through leaving them in their native form for statistical analysis.

Key Concepts: The Primary Classes of Data

- Categorical
  - Binary
  - Nominal
- Ordinal
- Numerical
  - Discrete
  - Continuous

Jargon Simplified: Rates and Proportions

Rates and proportions are distinct measures – often of disease – that are useful for describing and comparing data. Rates are a measure derived from taking the frequency or count of events divided by some measure of interval time. For example, the number of infections that occur in one year would be an infection rate. Proportions or percentages are unitless ratios of two quantities. This could be a quantity such as left ventricular ejection fraction or the number of patients out of a group of 100 that develop a complication after treatment with a certain drug. Understanding the difference between rates and proportions is crucial to their appropriate use as variables and outcome measures of disease.
Descriptive Statistics

Now that the basic forms of data have been introduced, one needs a way to begin to describe and summarize them. Descriptions of data generally aim to provide some information about the variability of data and can be done both graphically and quantitatively as an important early step in analysis. Let us first consider some common graphical representations. When categorical variables or events are counted, such as the number of hip fractures admitted to hospitals in a small town over a year’s time, the bar diagram is an excellent form of illustration (Fig. 42.1). For numerical data, “measures of central tendency” are used such as means (averages), medians, modes, and frequency distributions. Numerical data, such as the age of patients undergoing primary total hip arthroplasty, can be illustrated using another graphic called a histogram (Fig. 42.2).

The variability of data can be quantified in numerous ways. The simplest way to describe the spread of a dataset is to report the highest and lowest values (called the range) which in the example illustrated in Fig. 42.2 would be 41.3 years and 89.6 years, respectively. The range tells us about the most extreme values of a set of data, but very little about how those data are distributed. Centiles are a value below which a given percentage of the values occur (example 5 to 95th percentiles). One of the most common uses of centiles is the interquartile range (the difference between the 25th and 75th centiles). The interquartile range as well as the median, 2½ percentile and 97½ percentile values are nicely visualized using the semiquantitative box-and-whisker plot (Fig. 42.3).

Now let us consider more quantitative explanations of data. One of the most important concepts in showing the variability of data comes from the idea of averaging the distance each value is from the mean. A simple mean is calculated by adding up numerical values for a series of observations (for example, ages of 100 patients undergoing hip arthroplasty) and dividing this sum by the number of observations (in this case, 100). Given a mean (m), and an observed value (x), the distance of that observation from the mean is (m − x). If we sum the square power of the differences for (x), we get a positive number that can be averaged (by dividing by n − 1) and we obtain a quantity called the variance (V).

\[ V = \frac{\sum (m - x)^2}{n - 1} \]  
Eq. 42.1

The square root of the variance (V) is called the standard deviation (SD) and is a fundamental statistical quality of data. It is used in many more sophisticated statistical
sured in hours per week could be converted to minutes per week – a systematic change in our data points that would alter the slope, but have no effect on the relationship between the two variables.

So how do we measure the strength of a relationship between two continuous variables? It is actually quite simple. The closer the data points fall in a straight line to form a linear relationship, the stronger the relationship between the variables. Consider Fig. 44.3.

Looking at the two graphs in Fig. 44.3, it is apparent that both graphs represent a positive relationship between stress levels and hours worked. That is, increase in hours worked also leads to increases in stress levels. However, the points in Fig. 44.3a lie much closer to the fitted line, whereas those in Fig. 44.3b are scattered more widely above and below the fitted line. In Fig. 44.3a, the fitted line (which is a function of $x$) better captures the variance in $y$. In other words, by knowing the value of $x$ (hours worked), we are better able to predict the value of $y$ (stress level).

How exactly is the line of best fit created to capture the greatest amount of variance of the data points? The process is known as *linear regression analysis*.

### Linear Regression Analysis

Consider the graph below (Fig. 44.4). The data represents the resulting blood loss in femoral neck fracture patients treated with hemiarthroplasty.

The value of $y$ (blood loss) for any given value of $x$ (minutes) can be predicted by analyzing the fitted line. This predicted value of $\hat{y}$ (known as $y$-hat) can be explained by the basic formula for a straight line,

$$\hat{y} = mx + b$$  \hfill Eq. 44.1

**Fig. 44.3** (a) Sample graph of medical residents’ hours worked versus stress levels showing points lying closer to the fitted line, representing a stronger relationship between the two variables. (b) Sample graph of medical residents’ hours worked versus stress levels showing points scattered more widely around the fitted line, representing a weaker relationship between the two variables.

**Fig. 44.4** The graph plots operation length versus blood loss in femoral neck fracture patients treated with hemiarthroplasty and depicts the line of best fit.
where \( m \) is the slope of the fitted line and \( b \) is the y-intercept. However, as the graph depicts, all the points do not lie on the fitted line, and the above formula can only predict the value of blood loss with given error in most cases.

The true value of \( y \) is expressed by the following formula:

\[
y = mx + b + \varepsilon \tag{Eq. 44.2}
\]

The error term, \( \varepsilon \), explains the variance of the points from the fitted line due to factors, other than operation length, that influence blood loss. The error term is also known as the residual variance.\(^2\) The value of the error term for each point on the above graph is the vertical distance between the actual point and the fitted line, as spanned by the green line. Figure 44.4 demonstrates the error terms associated with the first three data points in Fig. 44.4 with a magnified view.

The goal of regression analysis is to find values for the slope and y-intercept that will yield a line of best fit that explains the variance in \( y \) to the highest degree possible. Essentially, we need a line that lies as close as possible to all the data points, so that the residual variance is minimized. This introduces the concept of linear regression analysis. Formally stated, linear regression analysis is the creation of a line of best fit that minimizes the sum of squares of all the error terms (sum of squares residual). This is also known as least-squares regression analysis. The product of this analysis, the fitted line, is known as the least-squares regression line.\(^1,2\) It is important to note that the error terms are not summed directly because some are positive values while others are negative, which will cancel each other if summed directly. Rather, regression analysis sums the square of each error term so that all values are positive.

**Jargon Simplified: Linear Regression Analysis and Least-Squares Regression Line**

Linear regression analysis: A statistical approach for measuring the linear relationship between two continu-

---

![Graph](image)

**Fig. 44.5** The graph plots operation length versus blood loss in femoral neck fracture patients treated with hemiarthroplasty and depicts the line of best fit.

---

**Coefficient of Determination and Correlation Coefficient**

Regression analysis provides values for the slope and y-intercept, which maximize the variance in \( y \) captured by the line of best fit. However, we have yet to discuss how to measure the magnitude of variance that is actually captured by this line. For that, it is essential to introduce the coefficient of determination. The coefficient of determination, expressed as \( r^2 \), measures the percentage of variation in \( y \) that is accounted for by the least-squares regression. It can also be viewed as the percentage of total variation in the dependent variable that is explained by the independent variable.\(^2\)

\[
r^2 = \frac{\text{variance of } \hat{y}}{\text{variance of } y} \tag{Eq. 44.4}
\]

or

\[
r^2 = \frac{\text{sum of squares regression}}{\text{sum of squares regression + sum of squares residual}} \tag{Eq. 44.5}
\]

The numerator in Eq. 44.4, the variance of \( \hat{y} \), is the sum of squares of \( \bar{y} - \hat{y} \) – the difference between the fitted point that corresponds to each of the actual data points and the mean value for the independent variable. This is a measure of the variance that the independent variable \( y \) would