REVIEW ARTICLE

Bright and dark sides of evidence-based medicine

Jacek Jassem

Department of Oncology and Radiotherapy, Medical University of Gdańsk, Gdańsk, Poland

KEY WORDS

evidence-based medicine, meta-analysis, publication bias, randomized clinical trial

ABSTRACT

Evidence-based medicine is a current paradigm of practicing and teaching clinical medicine. This concept is widely believed to optimize clinical decision making by using the evidence derived from high-quality medical research, in particular from randomized clinical trials (RCTs) and their meta-analyses, as they are associated with a lower risk of bias compared with observational studies. However, RCTs are not devoid of limitations and their results may be distorted for various reasons. This article reviews the most common shortcomings of RCTs, with particular focus on publication bias.

A habit of basing convictions upon evidence, and of giving to them only that degree or certainty which the evidence warrants, would, if it became general, cure most of the ills from which the world suffers. (Bertrand Russell, 1957)

Introduction Evidence-based medicine (EBM), defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients," is a contemporary paradigm of practicing and teaching clinical medicine.[1] This concept is based on the process of systematically reviewing, appraising, and using clinical research findings to aid the delivery of optimum clinical care to patients.[2] The practical methodology of EBM was developed in the late 1980s and early 1990s by Gordon Guyatt and David Sackett, clinical epidemiologists working at McMaster University in Hamilton, Ontario, Canada.[1,3] EBM has transformed medicine into a scientific discipline and is commonly regarded as the gold standard and the most important base for clinical practice. Somewhat humorously, the aims of EBM have been defined as the separation of evidence from propaganda, certainty from assumptions, study results from assertions, rationalism from superstitions, science from folklore, and knowledge from dogmas.

Evidence-based medicine as the basis for clinical decision making EBM is expected to optimize clinical decision making by using the evidence derived from high-quality medical research. However, research conclusions are made in a probabilistic manner. Hence, one cannot simply put them into "evidence-based" and "non–evidence-based" categories, as they are anywhere on a continuum. It is necessary, therefore, to define the risk of bias, that is, the level of certainty that the observed effect is true. There are several systems and permutations of evidence-certainty grading, most of which (with some simplifications) include 5 levels with decreasing confidence:

LEVEL 1 well-designed and executed randomized controlled trials (RCTs) and their systematic meta-analyses;
LEVEL 2 cohort studies and their systematic reviews, or low-quality RCTs;
LEVEL 3 case-control studies and their systematic reviews;
LEVEL 4 historical case series and their reviews;
LEVEL 5 case reports and testimonials.

This classification, in turn, is used for grading therapeutic guidelines: from the strongest (based on several high-quality RCTs with consistent results or, in special cases, on a single large high-quality RCT) to the weakest (lacking direct research evidence and based on empirical grounds). Implementing guidelines is expected to strengthen the scientific base for clinical decisions and to increase consistency, efficiency, quality, and safety of medical care.

Therapeutic guidelines are generally based on RCTs and their meta-analyses, as they are associated with a lower risk of bias compared with observational studies.

Correspondence to:
Prof. Jacek Jassem, MD, PhD, Katedra i Klinika Onkologii i Radioterapii, Gdański Uniwersytet Medyczny, ul. Dębinki 7, 80-211 Gdańsk, Poland, phone: +48 58 349 22 70, e-mail: jjassem@gumed.edu.pl
Received: March 27, 2016.
Accepted: March 27, 2016.
Published online: May 5, 2016.
Conflict of interest: none declared.
Pol Arch Med Wewn. 2016; 126 (5): 347-352
doi:10.20452/pamw.3403

Medicine: science or art? Despite their benefits, RCTs are not devoid of limitations. First, they may be subjected to flaws in their design and execution. For example, shortcomings of therapeutic RCTs include an insufficient number of patients, unbalanced distribution of prognostic factors in study arms, poor patient compliance, protocol violations, or concomitant use of other medications, to mention only a few. Additionally, methodology of RCTs may be irrelevant to clinical practice (eg, owing to very restrictive entry criteria). Finally, large RCTs, providing the strongest evidence, are expensive and difficult to run. This, in turn, may result in assigning the priority to research topics influenced by the sponsors' interests rather than to those considered most relevant by the medical community (in particular, pragmatic trials based on real practice). Indeed, around 80% of clinical trials are industry-sponsored because they are better funded and have easier design.[4]

The idea of EBM was accepted with great enthusiasm and hope, as it stressed a previously neglected role of research evidence. With time, however, there has been an increasing recognition that EBM should always be viewed as complementary to clinical experience. For example, therapeutic RCTs define whether one treatment is more efficacious than another on average, but by no means can they answer the question of how to treat an individual patient. Hence, the interpretation of clinical studies must be critical, including validity and importance of the results, as well as their relevance in clinical practice.[5] Judicious appraisal of EBM necessitates particular skills and knowledge, whereas clinicians often lack time, competence, and criticism to accomplish this requirement.[6] Applying EBM must also integrate best research evidence with availability and cost of treatment, as well as consider the patient's values, preferences, and expectations.[1,7,8] In other words, EBM should not be viewed as a "cookbook medicine", that is, it cannot consider all patients as essentially interchangeable. This debate over EBM is the return to the classic question of whether medicine is an art or a science.

An underestimated problem is the adherence of the RCT results to those seen in "real life" daily clinical practice. There is a large body of evidence showing that results from prospective clinical studies cannot be simply extrapolated into routine practice, owing to differences in patient selection and treatment compliance. Indeed, therapeutic effectiveness in routine clinical practice is usually lower than that reported in clinical studies, whereas the toxicity is higher.[9,10] The base for many therapeutic interventions is sparse or of poor quality, and clinical recommendations seem to be increasingly founded on lower levels of evidence.[11] There are numerous commonly accepted interventions, for example, insulin for diabetes, suturing for large wounds, or defibrillation for ventricular fibrillation, that have not been (and most likely never will be) addressed by RCTs.[12,13] However, missing evidence of efficacy does not imply the evidence of inefficacy. In some cases, such as emergency surgery, conducting RCTs would be unethical, still giving room for observational studies.

The good, the bad, and the ugly The application of EBM in the clinics may also be affected by faulty interpretation of otherwise properly conducted RCTs. A good illustration of this problem are RCTs that compared 2 postoperative endocrine therapies in breast cancer: tamoxifen, a well-known and nontoxic agent, and aromatase inhibitors. These studies, including tens of thousands of patients, almost uniformly demonstrated a small, yet significant, increase in the disease-free survival in favor of aromatase inhibitors, but none showed increased overall survival. Both therapies were usually associated with similar quality of life and different (although not favoring either) toxicity profiles. Notwithstanding, the results of these studies were notoriously presented in a way suggesting an apparent superiority of aromatase inhibitors.[14] Such interpretation has resulted in a massive and partially unjustified replacement of tamoxifen by aromatase inhibitors, and this trend has only recently been somewhat tempered. Another example of such misinterpretation in oncology is the study of comparing chemotherapy with or without erlotinib (an inhibitor of epidermal growth factor receptor) in advanced pancreatic cancer.[15] Although the erlotinib arm was more efficacious in statistical terms, the difference in median survival compared with the control arm was trivial (10 days) and achieved at the expense of increased toxicity and cost. Despite this, erlotinib was considered beneficial and found a wide application in clinical practice.

The biggest threat: publication bias An apparent limitation of many RCTs is the small number of patients, resulting in the low statistical power of comparisons. This weakness has partially been resolved by applying meta-analyses and systematic reviews, which increase statistical power through amalgamating, summarizing, and reviewing previous quantitative research. The limitations of meta-analyses include their retrospective design, not accounting for variations of standard treatments over time, and the heterogeneity of primary research. Most importantly, however, meta-analyses and systematic reviews may only be considered trustworthy if they include all relevant studies, whether published or not. Violation of this principle results in publication bias, a phenomenon defined at the end of the 20th century as the tendency for investigators to submit manuscripts and for journals to accept them, based on the strength and direction of the research findings.[16,17]

There are numerous examples demonstrating this problem. One of the most striking is the review of RCTs on antidepressant drugs, published in 2008.[18] Of the 74 studies registered by the US Food and Drug Administration (FDA),

31% (accounting for 3449 study participants) were not published. Of the 37 studies considered by the FDA as having positive results, 36 were published, as opposed to only 3 of the 36 studies viewed by the FDA as having negative or questionable results. Of the remaining 33 studies, 22 were not published and 11 were published in a way conveying a positive outcome. In consequence, based on the published literature, 94% of the trials conducted were positive, whereas actually 51% of the trials were viewed as positive by the FDA. Krzyżanowska et al[19] analyzed factors associated with failure to publish large RCTs presented at annual meetings of the American Society of Clinical Oncology (ASCO), the most prominent oncology conference worldwide. The inventory included all abstracts from large RCTs presented at ASCO meetings between 1989 and 1998. Of the 510 RCTs, 26% were not published in full within 5 years after presentation at the meeting, including 19% and 32% with "positive" and "negative" results, respectively (P <0.001). There was also a highly significant difference between median times to publication for studies with significant and nonsignificant results (2.2 and 3.0 years, respectively; P <0.001). Another example is even more astonishing. Among the 2028 oncology RCTs registered by the National Institutes of Health, only 17.6% were available in PubMed, including 59% and 5.9% of trials sponsored by academic institutions and by the industry, respectively.[20] A total of 65% of RCTs reported the results as positive, including 50% sponsored by academic institutions and 79% by the industry, suggesting a selective release of their research by pharmaceutical companies.

Growing competition for funding and citations has resulted in the increasing frequency of publications showing statistically significant, or otherwise "better", results (over 22% growth across disciplines between 1990 and 2007), and this trend was even stronger in biomedical sciences.[21] The most comprehensive analysis addressing this problem is the Cochrane review published in 2009.[22] The authors identified 5 studies that systematically investigated the extent to which the publication of clinical trials is influenced by the statistical significance or direction of a trial's results. The search included major publication databases: the Cochrane Methodology Register, MEDLINE, EMBASE, and Ovid, as well as the Science Citation Index. The findings of particular studies were classified as either positive (statistically significant at P <0.05, perceived as striking or important, or showing a positive direction of effect) or negative (not statistically significant, perceived as unimportant, or showing a negative or null direction of effect). Trials with positive findings had almost 4 times higher odds of being published compared with those with negative or null findings. Furthermore, trials with positive findings tended to be published after 4 to 5 years, compared with 6 to 8 years for those with negative findings.

There are a few potential reasons for publication bias. Firstly, sponsors may either not encourage or actively discourage investigators from publishing negative studies. Secondly, there is a limited enthusiasm of researchers to release studies deemed unworthy of publication. Finally, such studies are less likely to be accepted for publication, particularly if they contradict widely-held beliefs or notions. In consequence, negative studies remain increasingly hidden in file drawers.

Publication bias may carry important ethical and scientific implications. The former include patients' acceptance of participation in clinical trials in good faith and in the belief that they are contributing to the improving care of future patients. Further, the medical community expects a fair dissemination of research results. On scientific grounds, failing to publish a study would not be harmful, provided this would equally concern "positive" and "negative" studies, but this is apparently not the case. Such obvious distortion in making public "positive" and "negative" studies will inevitably result in dissemination of misleading medical knowledge, with all its clinical consequences.

Publication bias in situ An even more difficult to identify and largely ignored scientific threat to the integrity of the literature is the so called "publication bias in situ".[23] This bias is due to the fact that, irrespective of the study content, researchers are free to decide what to analyze, how to analyze, and what to report. Chasing for positive results includes extensive and unacceptable statistical manipulations ("data massage"), such as a post hoc selection of analyzed variables and analytical methods, multiple analyses from the same datasets, or attributing excessive prominence to subgroup analyses. This leads to picking up and publishing only selected, unplanned, but otherwise most appealing, aspects of particular studies. An example of such biased reporting is the US trial evaluating the efficacy of a human immunodeficiency virus (HIV) vaccine.[24] Study results were generally negative within the study population as a whole, but there was statistically significant reduction of HIV infection in certain vaccinated subgroups, constituting a small proportion of the entire series. Although this "positive" finding did not include correction for multiple testing and was not based on a prespecified hypothesis, the study was presented in the popular and business press as promising and attracted great attention.[25] The public appreciation of this notion was long maintained, even though the results of a subsequent study using the same vaccine were negative.[26]

Another example of a misleading message was the study comparing an aromatase inhibitor, letrozole, with tamoxifen in preoperative therapy of breast cancer.[27] This study enrolled 324 patients, including 39 with overexpression of type 1 or 2 epidermal growth factor receptor (HER1, HER2) in tumor cells. The first publication of this

study did not present the overall results; instead it emphasized (even in the title) a superiority of letrozole over tamoxifen in the small subset of HER1/2-positive patients. Although this finding was a result of a post hoc unplanned exploratory analysis and did not account for multiple comparisons, it was awarded a "rapid publication" status in a prestigious journal. In consequence, despite negative results of subsequent studies, this report created an unjustified tendency to a selective use of aromatase inhibitors in HER2-positive patients, also in other settings.

What is the way out? Regardless of its limitations, EBM remains the best available method of establishing the outcomes of medical procedures and is an indispensable part of modern medicine. However, the clinical research constituting its base may be subjected to important flaws and limitations. Whereas faulty clinical research may be reduced by refining its methodology and more stringent peer-review policies, there remains the problem of publication bias. Health science literature provides many examples of a tendency to show effects greater than the true value. These practices may produce highly confusing results and lead to dissemination of misguided policies. There is an apparent need for systematic efforts to identify, understand, and address this problem. A simple solution would be a generic rule of publishing all studies and all findings of particular studies, irrespective of their outcomes. This postulate could be easily accomplished in the era of online journals, but would require a broad consensus of the scientific community. Meanwhile, all efforts should be made to identify unpublished findings in order to include all relevant studies in reviews and meta-analyses.

Within the past years, there have been several attempts to increase the quality of clinical research and reduce the risk of publication bias. In 2005, the International Committee of Medical Journal Editors (ICNJE) declared that they would only consider a trial for publication if it has been registered before the enrollment of the first patient.[28] This requirement aimed at preventing selective publication and selective reporting of research outcomes. A strict endorsement and full implementation of this obligation by all medical journals will largely reduce the scale of publication bias. Most recently, the ICNJE proposed that authors should share with others the de-identified individual-patient data generated by interventional clinical trials, to allow reproducing the article's findings.[29] The data underlying the results should affect the fabric of how clinical trials are planned and conducted, and how their data are used. In 2014, *The Lancet* initiated a movement (*REduce research Waste And Reward Diligence*; REWARD), with the aim of increasing the quality of biomedical research and publishing all findings (http://www.thelancet.com/campaigns/efficiency). This initiative was based on the assumption that of a third of a trillion US dollars invested annually on biomedical studies across the world, a large part is wasted owing to their poor design, conduct, and reporting. The REWARD aims should be obtained through systematic and continuous monitoring of the performance of funders, regulators, academic institutions, journals, and researchers.

This article highlights selected controversies around EBM and indicates challenges in clinical applications of biomedical research data. Better understanding and addressing EBM shortcomings may optimize clinical decision making.

REFERENCES

1 Guyatt G, Cairns J, Churchill D, et al. Evidence based medicine. A new approach to teaching the practice of medicine. JAMA. 1992; 268: 2420-2425.

2 Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. BMJ. 1995; 310: 1122-1126.

3 Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ. 1996; 312: 71-72.

4 Brezis M. Big pharma and health care: unsolvable conflict of interests between private enterprise and public health. Isr J Psychiatry Relat Sci. 2008; 45: 83-89.

5 Practice Committee of American Society for Reproductive Medicine. Interpretation of clinical trial results. Fertil Steril. 2008; 90 (Suppl 5): S114-S120.

6 van Dijk N, Hooft L, Wieringa-de Waard M. What are the barriers to residents' practicing evidence-based medicine? A systematic review. Acad Med. 2010; 85: 1163-1170.

7 Greenhalgh T, Snow R, Ryan S, et al. Six 'biases' against patients and carers in evidence-based medicine. BMC Medicine. 2015; 13: 200.

8 Kelly MP, Heath I, Howick J, et al. The importance of values in evidence-based medicine. BMC Medical Ethics. 2015; 16: 69.

9 Sargent D. What constitutes reasonable evidence of efficacy and effectiveness to guide oncology treatment decisions? Oncologist. 2010; 15 (Suppl 1): S19-S23.

10 George SL. Reducing patient eligibility criteria in cancer clinical trials. J Clin Oncol. 1996; 14: 1364-1370.

11 Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA. 2009; 301: 831-841.

12 Glasziou P, Chalmers I, Rawlins M, et al. When are randomized trials unnecessary? Picking signal from noise. BMJ. 2007; 334: 349-351

13 Croft P, Malmivaara A, van Tulder M. The pros and cons of evidence--based medicine. 2011; 36: E1121-E1125.

14 Seruga B, Tannock IF. Up-front use of aromatase inhibitors as adjuvant therapy for breast cancer: The emperor has no clothes. J Clin Oncol. 2009; 27: 840-842.

15 Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol. 2007; 25: 1960-1966.

16 Chalmers I. Underreporting research is scientific misconduct. JAMA. 1990; 263: 1405-1408.

17 Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA. 1990; 206: 1385-1389.

18 Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008; 358: 252-260.

19 Krzyżanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. JAMA. 2003; 290: 495-501.

20 Ramsey S, Scoggins J. Commentary: practicing on the tip of an information iceberg? Evidence of underpublication of registered clinical trials in oncology. Oncologist. 2008: 13: 925-929.

21 Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012; 90: 891-904.

22 Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database of Systematic Reviews. 2009. Issue 1. Art. No.: MR000006.

23 Phillips CV. Publication bias in situ. BMC Med Res Methodol. 2004; 4: 20.

24 HIV gp120 vaccine – VaxGen: AIDSVAX, AIDSVAX B/B, AIDSVAX B/E, HIV gp120 vaccine – Genentech, HIV gp120 vaccine AIDSVAX – VaxGen, HIV vaccine AIDSVAX – VaxGen. Drugs R D. 2003; 4: 249-253.

25 Altman LK. Official hopes to explain AIDS vaccine disparities. New York Times: A24. February 25, 2003.

26 Pitisuttithum P, Gilbert P, Gurwith M, et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. J Infect Dis. 2006; 194: 1661-1671.

27 Ellis MJ, Coop A, Singh B, et al. Letrozole is more effective neoadjuvant endocrine ther-apy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial. J Clin Oncol. 2001; 19: 3808-3816.

28 De Angelis CD, Drazen JM, Frizelle FA, et al. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. Lancet. 2005; 365: 1827-1829.

29 Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data – a proposal from the International Committee of Medical Journal Editors. N Engl J Med. 2016; 374: 384-386.

ARTYKUŁ POGLĄDOWY

Blaski i cienie medycyny opartej na dowodach naukowych

Jacek Jassem

Katedra i Klinika Onkologii i Radioterapii, Gdański Uniwersytet Medyczny, Gdańsk

SŁOWA KLUCZOWE

badania kliniczne z randomizacją, medycyna oparta na dowodach naukowych, metaanaliza, stronnicze publikowanie wyników

STRESZCZENIE

Medycyna oparta na dowodach naukowych (*evidence-based medicine*) jest obecnie ogólnie przyjętym sposobem uprawiania i nauczania klinicznych dziedzin medycyny. Powszechnie uważa się, że koncepcja ta zapewnia najwyższy poziom podejmowania decyzji klinicznych, ponieważ opiera się na dowodach z badań naukowych o wysokiej jakości – szczególnie na dowodach z badań klinicznych z randomizacją (*randomized clinical trials* – RCT) i ich metaanaliz, bowiem ryzyko popełnienia błędnej oceny jest w nich mniejsze niż w badaniach obserwacyjnych. Z drugiej jednak strony RCT zawierają pewne ograniczenia, a ich wyniki mogą zostać z różnych powodów zniekształcone. W artykule przedstawiono najważniejsze niedoskonałości badań klinicznych z randomizacją, ze zwróceniem szczególnej uwagi na zjawisko stronniczego publikowania ich wyników (*publication bias*).

Adres do korespondencji:
prof. dr hab. med. Jacek Jassem, Katedra i Klinika Onkologii i Radioterapii, Gdański Uniwersytet Medyczny, ul Dębinki 7, 80-211 Gdańsk, tel.: 58 349 22 70, e-mail: jjassem@gumed.edu.pl
Praca wpłynęła: 27.03.2016.
Przyjęta do druku: 27.03.2016.
Publikacja *online*: 05.05.2016.
Nie zgłoszono sprzeczności interesów.
Pol Arch Med Wewn. 2016; 126 (5): 347-352
doi:10.20452/pamw.3403