

Medical Students Can Help Avoid the Expert Bias in Medicine

Franz Porzsolt^{1*}, Peter Braubach¹, Petra Inge Flurschütz¹, Alex Göller¹,
Maria Barbara Sailer¹, Manfred Weiss², Peter Wyer³

¹Clinical Economics at the Institute of History, Philosophy and Ethics in Medicine,
University of Ulm, Ulm, Germany

²Department of Anesthesiology, University Hospital, Ulm, Germany

³Associate Clinical Professor of Medicine, Columbia University College of Physicians and Surgeons,
New York, USA

Email: *franz.porzsoelt@uniklinik-ulm.de, peter.braubach@uni-ulm.de, petraflurschuetz@web.de,
alex.goeller@uni-ulm.de, maria.sailer@yahoo.de, manfred.weiss@uniklinik-ulm.de, pw91@columbia.edu

Received August 23rd, 2012; revised September 21st, 2012; accepted October 3rd, 2012

Background: Applying the principles of Evidence Based Health Care (EBHC) in an academic environment we became aware of important differences between medical students and the users of clinical research. The latter may be clinicians, educators, guideline developers, or industry managers. These users are adapted to the system and have some kind of conflict of interest: they are either biased by patients' demands, by main stream thinking, by medical standards and/or by economic interests. All are under time pressure, want to avoid conflicts with their employer and the analysis of scientific reports may not really be their main focus. Medical students were not exposed to these confounders. **Methods:** Medical students under professional supervision completed critical assessments of more than 100 published studies. Their analyses were limited to identification of mistakes, bias and errors using a check list of potential weaknesses in design and conduct but included the feedback to the academic supervisors. **Results:** Medical students trained in Evidence Based Medicine are capable of identifying problems in clinical trials by the systematic application of an assessment checklist. **Conclusion:** In our approach we demonstrate that students can assist health care professionals and academic teachers with the assessment of clinical evidence. The premise of the approach is that the final appraisals, which involve consideration of clinical, practical and value issues, necessarily reside with the academic teachers, writers of guidelines or industry managers who constitute the active users of research.

Keywords: Evidence Based Health Care; Medical Education; Validity of Scientific Evidence

Introduction

Critical appraisal of scientific literature is an essential step when applying the principles of evidence-based medicine (EBM) (Guyatt et al., 1992). EBM may have different meanings to different people starting at a decision for a single patient and extending up to the definition of a complete guideline. If the principles of scientific evidence are scrupulously applied, a critical appraisal is time consuming. Especially the analysis of studies published in recognized journals often leads to discussions lasting several hours (Porzsolt et al., 2009).

The time and logistical demands of rigorous critical appraisal reflects on multiple levels of application. Many, if not most, protocols used in the development of practice guidelines for wide spread use restrict the evaluation of evidence supporting the recommendations to categorization of study designs. Only recently have systems emerged that take into account also the quality of individual studies (Schünemann et al., 2006) as elaborated in the PRISMA statement (Liberati et al., 2009). It will be difficult to detect bias if the critical review of evidence in guidelines is restricted to the confirmation of the levels of evidence (Kopp et al., 2008). Conscientious readers of such guidelines can tell whether the described validity of a study or whether the level of evidence merely describes the kind of in-

formation source (e.g., meta-analysis, cohort study, expert opinion).

The feasibility of practicing physicians performing viable critical appraisal on their own has been questioned in the literature (Guyatt et al., 2000). Practitioners aspiring to evidence-based care may be best advised to rely on publications such as ACP Journal Club, in which professionally done critical appraisals of leading clinical studies are performed and supervised by epidemiologists trained in the principles of EBHC but not involved in the clinical practice areas relevant to the included studies. There is abundant literature on teaching principles of critical appraisal to undergraduates (Norman et al., 1998). Although such instruction has been shown to increase knowledge and skills, it has not tended to reflect in changes in clinical behavior (Coomarasamy et al., 2004). Survey information indicates that medical undergraduates perceive that their limited ability to actively apply principles of EBHC impedes acceptance of their importance (Ilic et al., 2010). We perceived an opportunity to develop a unique framework for application of medical students critical appraisal and EBHC skills by turning their relatively undeveloped clinical skills into an asset, at the same time providing them with an important role to play within the 'food chain' that leads from the production of clinical research to its consumption within the various venues of clinical policy and practice.

*Corresponding author.

The aim of our study was to propose a method which increases the chance to detect mistakes and biases in scientific publications. The method should be easy to perform, inexpensive, fast and acceptable to the users of EBM.

Methods

About three years ago we started to invite medical students who knew already the medical discipline of their future specialization to complete a medical thesis on the “usability of scientific publications (USP)”. These students should ask a mentor of this discipline to suggest a topic which is discussed controversially. The students were trained in an EBM weekend course and received a structured questionnaire (**Table 1**) which they had to use for preparation of their assessments of the selected topic. The completed assessments and the assessed publications were distributed to the participants of the USP discussion group. In an average session of this group, 2 - 3 papers were discussed. The average group consisted out of two students and two supervisors. All members of the group had to read the paper and to prepare their comments. One of the students had to present the paper and his critical comments which was discussed and supplemented by the other three members of the group. One of the supervisors was a clinical expert in the area of the discussed paper, the second supervisor was also clinician with experience in clinical epidemiology. The students were motivated to participate as they could either complete a doctoral thesis (which is in Germany usually completed during the last years at medical school) or received payment which was slightly higher than the payment as service in pubs or academic student tutorship. The duration of a session was usually two hours. Some of the papers had to be discussed in more than one session because additional background information was needed.

The two examples presented in this paper (Beckett et al., 2008; Sprung et al., 2008) were derived from projects of our students (AG, PIF, MBS) who did critical appraisals on the

validity of about 100 publications on lipolysis (AG), on recommendations in intensive care units (PIF) and in trauma surgery (MBS). The results were used to discuss the validity of the analyzed studies and finally a consensus statement was created. This assessment is only the first step of a complete appraisal. The second step, the appraisal itself, is left to the decision maker. We identified four conditions that must be fulfilled to complete a reliable assessment of the validity of a scientific publication.

- Persons must be available who can perform an appraisal;
- The effort required to train and supervise these persons should be as little as possible;
- The persons who are to perform the critical appraisal must be motivated to do it;
- The professional authorities for whom these critical appraisals are completed must not be challenged.

These four prerequisites are met by medical students if they are offered introductory seminars, regular supervision and appropriate motivation. The seminars focus on the basics of EBM and practical experience applying its specific assessment methods to published papers. The papers selected for training contain frequently seen errors, such as study design which is inappropriate to answer the study question, inadequate descriptions of risk factors, and differences between the number of participating patients and those reported in the final evaluations. Most of these errors are not difficult to detect but the detection is time consuming and time is a permanent problem to any clinician.

The students had to prepare a written version of their assessments using a standardized questionnaire (**Table 1**). These assessments were compared, discussed, and a consensus protocol was edited during the group meetings.

The assessment starts by identifying the scientific question investigated in the study. The appropriate study question is usually extracted from the end of the introduction section. The students have then to construct an “ideal” and rather simple study design which could be used to answer the study question. This “ideal” study design must be compared with the study design used in the publication analyzed.

The students then assess whether the baseline criteria of the included study populations are comparable and if: all obvious risk factors have been accounted for, blinding was maintained, follow-up was long enough, and all patients who were included in the study were also included in the calculation of the results. If questions arise which cannot be solved by the participants of the discussion groups, additional experts (statisticians, clinical experts) are consulted. The final report containing a short summary can be used as a brief and lucid statement on the method used in the trial.

Results

The questions presented in **Table 1** were slightly modified several times. In Question #3 it should be stated whether or not the same study design will be in a second study in which the results of the first study had to be confirmed. Question #4 was phrased in a rather broad meaning as there were too many possibilities for inadequate conduct. Question #8 addressed the point that some of the studies were published as “blinded” although the results strongly suggested that doctors and/or patients could identify rather fast the allocation. Question #11 turned out to be less important as the use of questionable statis-

Table 1.
Standardized questionnaire used to assess the validity of medical publications.

1	Study question clearly stated?
2	Design appropriate to answer study question?
3	Can the study design be used in a confirmatory study?
4	Study conducted properly?
5	Appropriate endpoints selected to answer study question?
6	Risk profiles of the study populations similar?
7	Allocation to study groups concealed?
8	Doctors and patients continuously blinded?
9	Follow-up long enough to detect endpoints?
10	All patients included in reported results?
11	Adequate statistics applied?
12	Results possibly influenced by conflicting interests?
13	Validity of report acceptable?
14	Described effect clinically relevant?

tical methods was usually associated with many other flaws. On the other hand we called a statistician in the rare cases where we were concerned about the statistics despite in an otherwise carefully designed and completed study. By answering the last questions the students should express their impression about a possible conflict of interest (Question #12) sponsoring, authorship and message of the publication, whether or not the study really demonstrated (Question #13) what it claimed to demonstrate and whether or not the demonstrated effect was considered clinically important (Question #14).

The evaluations of the two studies by Sprung et al. (2008) and Beckett et al. (2008) are shown as table which represents the final results of our assessment (**Table 2**). The abstract of the paper by Sprung et al. (2007) was used as reference in a guideline for treatment of patients with severe sepsis and septic shock (Dellinger et al., 2008). Our second example (Beckett et al., 2008) was used for teaching purposes and was listed among significant publications of cardiovascular literature in 2008.

The problems we identified in the assessed studies were related to the inclusion of patients who met the exclusion criteria as described in the methods section of the publication (Sprung et al., 2008), to imbalanced risk profiles of the included patients and to inclusion of fewer patients than requested according to the power calculation of the study. In the second example (Beckett et al., 2008) the study question was not clearly stated, the study design was not appropriate, there were too many con-

founders, the study was not conducted properly, the risk profiles in the compared groups were not well balanced neither were all patients included in the reported results (**Table 2**).

The assessment of the study by Beckett et al. (2008) was supplemented by a detailed description of the study method and study conduct.

- It was not clear to the assessors how many patients were recruited for the run-in-phase;
- The criteria for stratification into the 4 study groups changed during the study. Between 2000 and 2003 a diastolic blood pressure of 90 - 109 mmHg was required. After 2003 a diastolic blood pressure of <110 and systolic blood pressure (standing) of >139 was introduced into the study protocol;
- The investigator could administer a second treatment to patients that did not reach an acceptable blood pressure with only the study medication. This intervention was again placebo controlled.
- Patients were withdrawn from the study if the additional treatment failed.
- Data from more than 300 patients was censored from the study some because the local investigator died or retired.
- Some data was analyzed according to the intention to treat principle while other was analyzed per protocol. The assessors did not find a proper reasoning for this kind of data analysis.

The assessment was concluded with a statement that the

Table 2.
Results of the assessment of two studies^{10,11} as presented in our project.

#		Sprung et al. ¹¹	Beckett et al. ¹⁰
1	Study question clearly stated?	Yes. (...), evaluated the efficacy and safety of low-dose hydrocortisone therapy in a broad population of patients with septic shock (...).	Unclear, see page 1888/para 3 left/lines 8 - 12
2	Design appropriate to answer study question?	Yes	No. See Appendix 1
3	Can the study design be used in a confirmatory study?	Yes	No. Too many confounders
4	Study conducted properly?	No. Patients with exclusion criteria were included. Study was underpowered.	No. See Appendix 1
5	Appropriate endpoints selected to answer study question?	Yes	Yes
6	Risk profiles of the study populations similar?	No. In the placebo arm there were more patients with severe diseases.	No. Risk marginally higher in six risk indicators in the placebo group.
7	Allocation to study groups concealed?	Yes	Unclear
8	Doctors and patients continuously blinded?	Yes	No. See protocol
9	Follow-up long enough to detect endpoints?	Yes. 28 days follow-up is acceptable in ICU studies.	Yes
10	All patients included in reported results?	Yes	No. See Figure 1 , page 1891.
11	Adequate statistics applied?	Yes	Yes
12	Results influenced by conflicting interests?	No	Possibly
13	Validity of report acceptable?	No. Study included less than the recommended number of 800 patients.	No. Too many confounders
14	Described effect clinically relevant?	No. Due to lack of validity	No. Due to lack of validity

complex study protocol may have given way to many possible confounders. Overall there are many problems concerning the validity of the study severely limiting the possibility to derive meaningful conclusions for clinical practice.

Discussion

It is generally recommended to check carefully the quality of scientific information especially if used not only for individual patient decisions but rather to support students' teaching, the writing of guidelines or the approval of new drugs (Ford et al., 2010). Both of the papers analyzed here were critically reviewed in the ACP journal club and were at the end accepted by clinical commentators as relevant to clinical decision making (Hung et al., 2008; Lamontagne et al., 2008). Our reviews may appear to have come to different conclusions. None of these two papers were considered to be devoid of potentially important flaws. It has to be considered that there are four essential differences between the ACP reviews and our reviews.

- We address specified validity criteria and intent to be as critical as possible to avoid mistakes, bias, and errors but not providing an assessment such as those provided by the clinical commentators of ACP Journal Club synopses.
- The appraisal is left to the persons who are responsible for the use of the published information. The ACP reviews provide assessments which are accompanied by an appraisal and this appraisal can directly be used by the reader of the ACP review.
- Our assessments are usually not prepared for publication. The publication of the two above assessments is necessary to provide examples. If this rigorous approach is published it will create a serious problem because economic competitors, i.e. other pharmaceutical companies may use the crude information for marketing purposes. The emerging ethical conflict in the cooperation of researchers from industry and academy was recently addressed in a comment (Fangerau, 2009).
- Finally, our assessments include two feed-back modules which are essential in teaching (Hattie, 2009). One is the didactic feed-back. The second is an epistemological feed-back between students and supervisors—and this is sometime rather time-consuming—because discussions about truth, believe, justification based on different degrees of skepticism cannot be avoided. This complete circle of feed-back is a basic requirement for “reframing” i.e. for generating consensus in a situation where the discussants do not agree.

Of the two papers illustrated in the example, one of them was used as a reference in a guideline (Sprung et al., 2008), and the other paper was used for teaching purposes and was considered the leading article in a list of significant publications in the cardiovascular literature in 2009 (Becket et al., 2008).

Our results indicate that medical students can be quickly trained in the required attitudes, skills, and knowledge of EBM to complete valuable critical appraisals (Finkel et al., 2003). Students can be motivated to do this if they can generate incomes comparable to other student jobs. Finally, acrimony among professional authorities can be avoided if the process of critical appraisal is divided into two parts. The students exclusively perform assessment and present their completed assessments with their evaluative statement only to the client who requested the assessment. The appraisal, including acceptance of the final

result of the entire process, has to be made by the clients, i.e. professional academic teacher, the professional researcher, or the writer of guidelines. The final decisions are left to the health-care professionals. Students lack clinical and research experience as well as measures of appropriateness. In addition, experts and students have different expectations in the validity of scientific information which in fact is a matter of judgment (Rawlins, 2008; Schulz et al., 2010).

We are convinced that most of the 40 available instruments for critical appraisals (Darmoni et al., 2001; Forestier et al., 2005) can be used for such a project. The result of the project will probably depend on the depth of the critical discussion.

It is essential that the students who use the questionnaire for validity assessment are aware of the most frequent mistakes and biases and know the questions that have to be asked for a profound assessment of the validity. In **Table 3** we demonstrate that our questionnaire and the instrument recommended by CASP, the Birmingham Critical Appraisal Skills Programme (CASPb, 2012) is similar, except that the CASP instrument includes a question about randomization and the assessment of the study results. The CASP instrument places less emphasis on the details of the study design, on the follow-up time, and on a possible conflict of interest. Our instrument neither addresses the question of whether the study results would apply to the local population nor investigates the precision of the data. Our results indicate that even the concise version of a questionnaire is sufficient to detect mistakes and bias if its users are trained to recognize the problems that may be hidden in a study with brilliant results.

Another instrument for a systematic appraisal is the GRADE system (Guyatt et al., 2008) recommended for the preparation of clinical guidelines. Although different versions of GRADE are not absolutely identical (**Table 4**) they cover almost the same questions as our instrument and the CASP instrument as far as these questions are related to RCTs.

We believe that students can make significant contributions at different stages in the generation and application of scientific evidence (**Figure 1**) and may therefore have a direct effect on the experts' judgment which is an essential ingredient of decision making (Rawlins, 2008).

This method may also be useful as a teaching tool for EBM as requested by Murad et al. (2009). The detailed discussion of the study methodology mediates the understanding of EBM (Plint et al., 2006) without giving students the impression of attending a theoretical seminar without significance to their future job. The students' appraisal cannot have the function of a final scientific appraisal but may induce in the client considerations about the acceptability of the evidence. In addition, students who are not native English speakers are supported to overcome their well known language problems (Letelier et al., 2007). This expected benefit of our approach could meanwhile be confirmed in cooperation with an Italian Group (Rosati et al., 2009; Rosati et al., 2012).

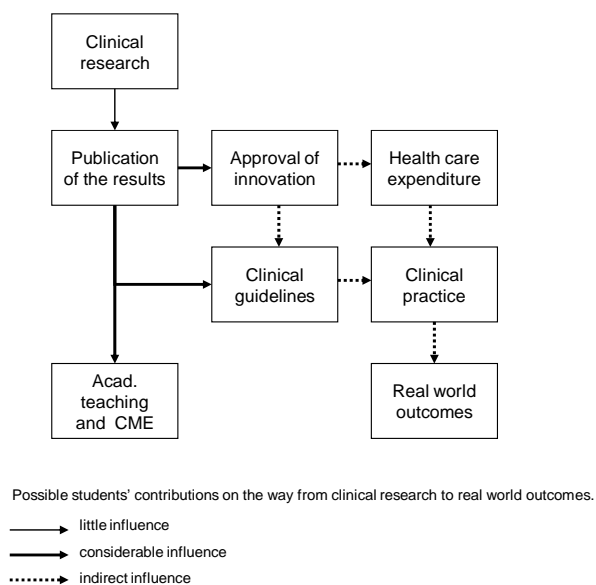
In summary, the students provide a critical appraisal of information which has been accepted by experts and is published. The information is provided within a short time at low expense. It is checked by an academic supervisor, but leaves the appraisal of the evidence and, therefore, the final decision about acceptance or rejection of the evidence to the person who requested the assessment. Finally, the students and the institution which organized their supervision will benefit from this program, as the students acquire valuable experience they can use

Table 3.
Comparison of the CASP questionnaire and the questionnaire used in this project.

#	CASP questionnaire	#	Our questionnaire
A1	Did the trial address a clearly-focused issue? An issue can be "focused" in terms of - the population studied - the intervention given - the outcomes considered	1	Study question clearly stated?
		2	Design appropriate to answer study question?
A2	Was the assignment of patients to treatments randomized?		
A3	Were all of the patients who entered the trial properly accounted for at its conclusion?	10	All patients included in reported results?
		7	Allocation to study groups concealed?
B4	Were patients, health workers, and study personnel "blind" to treatment? - were the patients - were the health workers - were the study personnel	8	Doctors and patients continuously blinded?
B5	Were the groups similar at the start of the trial? <i>In terms of other factors that might effect the outcome, such as age, sex, social class</i>	6	Risk profiles of the study populations similar?
B6	Aside from the experimental intervention, were the groups treated equally?	2	Design appropriate to answer study question?
		4	Study conducted properly?
B7	How large was the treatment effect? <i>What outcomes were measured?</i>	11	Described effect clinically relevant?
B8	How precise was the estimate of the treatment effect? <i>What are its confidence intervals limits?</i>		
C9	Can the results be applied to the local population? <i>Do you think that the patients covered by the trial are similar enough to your population?</i>	3	Can the study design be used in a confirmatory study?
C10	Were all clinically important outcomes considered? <i>If not, does this affect the decision?</i>	5	Appropriate endpoints selected to answer study question?
C11	Do the benefits outweigh the harms and costs? <i>This was probably not addressed by the trial, but what do you think?</i>	11	Described effect clinically relevant?

Table 4.
Quality criteria incorporated into GRADE system.

Randomized Trials (ref 26)	Allocation concealment
	Lack of blinding, including outcome adjudicators-more important when outcome assessment potentially subjective and subject to bias
	Lack of intention to treat analysis
	Stopping early for benefit
	Failure to report important outcomes
Randomized Trials (ref 25)	Use of unvalidated patient reported outcomes
	Carry over effects in cross over trials
	Recruitment bias in cluster randomized trials
Observational Studies (ref 26)	Removal of all biases would increase effect size
Observational Studies (ref 25)	Under or over-matching of case-control studies
	Selection of exposed and unexposed in cohort studies from different populations
	Differences in measurement of exposure (e.g. recall bias in case-control studies)
	Differential surveillance for outcome in exposed and unexposed in cohort studies
	Failure of accurate measurement of all known prognostic factors
	Failure to match for prognostic factors and/or adjustment in statistical analysis
General	Depending on the context and study type, there can be additional limitations than those listed above. Guideline panels and authors of systematic reviews should consider all possible limitations.

**Figure 1.**

Possible students' contributions on the way from clinical research to real world outcomes. —> little influence, → considerable influence,> indirect influence.

in their later professional life. Institutions that can continuously guarantee a high-quality output will contribute to an urgently need change: we will have to become be much more critical in what we accept if our primary goal is to maintain the best possible but also affordable health care system.

REFERENCES

- Beckett, N. S., Peters, R., Fletcher, A. E. et al. (2008). Treatment of hypertension in patients 80 years of age or older. *Journal of Medicine*, 358, 1887-1898. doi:10.1056/NEJMoa0801369
- CASPb Working Group (2012) Crib sheet for an RTC: 11 questions to help you make sense of a trial. URL (last checked 29 January 2012). <http://medweb4.bham.ac.uk/websites/caspb/cribsheets>
- Coomarasamy, A., & Khan, K. S. (2004). What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *British Medical Journal*, 329, 1017. doi:10.1136/bmj.329.7473.1017
- Darmoni, S. J., Haugh, M. C., Lukacs, B., & Boissel, J. P. (2001). Quality of health information about depression on internet. Level of evidence should be gold standard. *British Medical Journal*, 322, 1367.
- Dellinger, R. P., Levy, M. M., Carlet, J. M. et al. (2008). Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Medicine*, 34, 17-60. doi:10.1007/s00134-007-0934-2
- Fangerau, H. (2009). Research ethics, publication ethics and the dialectics of scientists trying not to behave badly: A comment on the advantages and limitations of twin assessment of clinical trials (TACT). *Journal of Public Health*, 17, 433-434. doi:10.1007/s10389-009-0284-3
- Finkel, M. L., Brown, H. A., Gerber, L. M., & Supino, P. G. (2003). Teaching evidence-based medicine to medical students. *Medical Teacher*, 25, 202-204. doi:10.1080/0142159031000092634
- Ford, A. C., Guyatt, G. H., Talley, N. J., & Moayyedi, P. (2010). Errors in the conduct of systematic reviews of pharmacological interventions for irritable bowel syndrome. *American Journal of Gastroenterology*, 105, 280-288. doi:10.1038/ajg.2009.658
- Forestier, R., Françon, A., & Graber-Duvernay, B. (2005). Les paramètres de validité d'un essai thérapeutique et leur influence sur l'élaboration d'une médecine fondée sur les preuves: revue de la littérature. *Annales de Réadaptation et de Médecine Physique*, 48, 250-258. doi:10.1016/j.annrmp.2005.02.004
- GRADEpro Software Package (2012). URL (last checked 29 January 2012). <http://ims.cochrane.org/revman>.
- Guyatt, G., Cairns, J., Churchill, D. et al. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, 268, 2420-2425. doi:10.1001/jama.1992.03490170092032
- Guyatt, G., Meade, M., Jaeschke, R., Cook, D., & Haynes R. B. (2000). Practitioners of evidence based care. *British Medical Journal*, 320, 954-955. doi:10.1136/bmj.320.7240.954
- Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schunemann, H. J. (2008). What is "quality of evidence" and why is it important to clinicians? *British Medical Journal*, 336, 995-998. doi:10.1136/bmj.39490.551019.BE
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hung, W. W., & Leipzig, R. M. (2008). Antihypertensive therapy with indapamide and perindopril reduced mortality in patients greater than or equal to 80 years. *ACP Journal Club*, 149, 10.
- Ilic, D., & Forbes, K. (2010). Undergraduate medical student perceptions and use of Evidence Based Medicine: A qualitative study. *BMC Medical Education*, 10, 58. doi:10.1186/1472-6920-10-58
- Kopp, I., Thole, H. T. L., Selbmann, H., & Ollenschläger, G. (2008). Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI): Fassung 2005/2006. *ZaeFQ*, 99, 468-519.
- Lamontagne, F., & Meade, M. O. (2008). Low-dose hydrocortisone did not improve survival in patients with septic shock but reversed shock earlier. *ACP Journal Club*, 148, 6.
- Letelier, L. M., Zamarin, N., Andrade, M. et al. (2007). Exploring language barriers to Evidence-based Health Care (EBHC) in post-graduate medical students: A randomised trial. *Education Health*, 20, 82.
- Liberati, A., Altman, D., Tetzlaff, J. et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *British Medical Journal*, 339, b2700. doi:10.1136/bmj.b2700
- Murad, M., Montori, V., Kunz, R. et al. (2009). How to teach evidence-based medicine to teachers: Reflections from a workshop experience. *Journal of Evaluation Clinical Practice*, 15, 1205-1207. doi:10.1111/j.1365-2753.2009.01344.x
- Norman, G. R., & Shannon, S. I. (1998). Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: A critical appraisal. *Canadian Medical Association Journal*, 158, 177.
- Plint, A. C., Moher, D., Morrison, A. et al. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *The Medical Journal of Australia*, 185, 263.
- Porzsolt, F., Bonotto, de O., Costa, I. C., & Thomaz, T. G. (2009). Advantages and limitations of Twin Assessment of Clinical Trials (TACT). *Journal of Public Health*, 17, 425-435. doi:10.1007/s10389-009-0283-4
- Rawlins, M. (2008). De testimonio: On the evidence for decisions about the use of therapeutic interventions. *Clinical Medicine, Journal of the Royal College of Physicians*, 8, 579-588.
- Rosati, P., Ciampalini, P., Grossi, A., Giovannelli, L., Giustini, F., Inglese, R., Fiscarelli, E., Castellano, C., Mazziotta, M. R. M., Gentile, S., Giampaolo, R., & Porzsolt, F. (2009). An alternate evaluation of evidence: Results from a CASP workshop. In P. J. Sharek, D. Bergman, & F. M. Ducharme (Eds.), *Beclomethasone for asthma in children: Effects on linear growth* (pp. 21-22). London: Cochrane Airways Group. doi:10.1002/14651858.CD001282
- Rosati, P., & Porzsolt, F. (2012). A practical educational tool for teaching child-care hospital professionals attending evidence-based practice courses for continuing medical education to appraise internal validity in systematic reviews. *Journal of Evaluation Clinical Practice*.
- Schünemann, H. J., Fretheim, A., & Oxman, A. D. (2006). Improving the use of research evidence in guideline development: 9 grading evidence and recommendations. *Health Research Policy and Systems*, 4, 21. doi:10.1186/1478-4505-4-21
- Schulz, K., Altman, D., & Moher, D. (2010). CONSORT 2010 state-

ment: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, 8, 18. [doi:10.1186/1741-7015-8-18](https://doi.org/10.1186/1741-7015-8-18)
Sprung, C. L., Annane, D., & Briegel, J. (2007). Corticosteroid therapy of septic shock (CORTICUS). *Abstract American Journal of Respiratory and Critical Care Medicine*, 175, A507.

Sprung, C. L., Annane, D., Keh, D. et al. (2008). Hydrocortisone therapy for patients with septic shock. *The New England Journal of Medicine*, 358, 111-124. [doi:10.1056/NEJMoa071366](https://doi.org/10.1056/NEJMoa071366)