

Critical appraisal of RCTs by 3rd year undergraduates after short courses in EBM compared to expert appraisal

Abstract

Introduction: An essential aim of courses in evidence-based medicine (EBM) is to improve the skills for reading and interpreting medical literature adequately. Regarding the conceptual framework, it is important to consider different educational levels.

Aim: Our primary aim was to investigate the applicability of different instruments for the assessment of methodological study quality by 3rd grade students after short courses in EBM. Our secondary outcomes were agreement with expert assessments and student's knowledge and competences.

Methods: We conducted four short courses in EBM of 90 minutes each for health care management and medical students focused on critical appraisal of the literature. At the end, the students assessed five publications about randomized controlled trials (RCTs) using five different instruments; the results were compared to expert assessments.

Results: In total, 167 students participated in our EBM courses. Students' assessments showed a non-systematic over- and underestimation of risk of bias compared to expert assessments with no clear direction. Agreement with expert assessments ranged between 66% to over 80%. Across RCTs, evidence was found that the choice of instrument had an impact on agreement rates between expert and student assessments ($p=0.0158$). Three RCTs showed an influence of the instrument on the agreement rate ($p<0.05$ each).

Discussion: Our results contrast sharply with those of many other comparable evaluations. Reasons may be a lack of students' motivation due to the compulsory courses, and the comparison to a reference standard in addition to self-ratings causing objectivity.

Conclusion: Undergraduates should become familiar with the principles of EBM, including research methods, and the reading of scientific papers as soon as possible. For a deeper understanding, clinical experience seems to be an indispensable precondition. Based on our results, we would recommend an integration of lectures about EBM and critical appraisal at least twice during studies and with greater intensity shortly before graduation.

Keywords: Critical appraisal, evidence-based medicine, randomized controlled trial, training, undergraduates

B. Buchberger¹

J.T. Mattivi¹

C. Schwenke²

C. Katzer¹

H. Huppertz¹

J. Wasem¹

1 University of Duisburg-Essen, Faculty of Economics and Business Administration, Institute for Health Care Management and Research, Essen, Germany

2 SCO:SSiS, Schwenke Consulting: Strategies and Solutions in Statistics, Berlin, Germany

1. Introduction

The level of awareness towards evidence-based medicine (EBM) is growing worldwide and the acceptance of its concept is increasing. In January 2007 the British Medical Journal conducted an online poll about the 15 most important medical milestones and EBM was in seventh place, right behind germ theory and oral contraceptive pill but ahead of computer and medical imaging [1], [2]. However, fostering an EBM culture and implementing it into practice requires the skills for identifying and appraising the literature critically [3], [4], [5]. A certain knowledge of probability and statistics is mandatory as well when accessing guidelines and evidence summaries, assessing

marketing and advertising material from industry, interpreting the results of a screening test, or reading research publications for staying up to date with newly developed treatments; furthermore, knowledge of biostatistics is necessary for the analysis of numerical data, for informing patients about treatment risks, and last but not least for being prepared to the Internet-literature of varying quality presented by patients [6], [7]. Actually, the question is no longer whether to teach EBM but how to teach it [8] and when. Apart from various educational methods, e.g. on the job training, problem-based or self-directed learning [9], the EBM concept may be taught as a whole as well as some of the five steps separately, which are

1. asking a clinical question,
2. searching for the best evidence,
3. critical appraisal of the evidence,
4. applying evidence to practice,
5. self-assessment [8].

There are quite a few courses introducing into database searches supported by librarians [9], or clubs as a format of training critical appraisal of the literature [9], [10], [11]. For measuring the increase of learners' competency by attending lectures in EBM, objective measurements are required rather than self-ratings leading to considerable overestimation [12], [13].

Previous studies evaluated the impact of EBM lectures mostly by self-reports of participants [4], [11] or kinds of question papers with multiple-choice questions [14]. In addition to self-assessments, the current study aims to achieve a certain objectivity by comparing students' assessments with a reference standard created by expert assessments.

Our primary outcome was to investigate the applicability of different instruments for the assessment of methodological study quality by 3rd grade students after short courses in EBM. Our secondary outcomes were agreement with expert assessments and student's knowledge and competences.

2. Methods

We included medical students directly after the preclinical phase at the faculty of medicine and students of the master's program "health care management" at the faculty of economic sciences of the University of Duisburg-Essen, Germany. They were trained in the principles of EBM in four sessions of 90 minutes each which were held in the context of lectures on health economics (compulsory courses for medical students) by the Institute of Health Care Management and Research in the winter semester of 2013/2014. The number of participants was limited to 20 students per session. A glossary of terms for quick searches was handed out (see Attachment 1).

2.1. Description of instruments and experts

Based on a Health Technology Assessment Report, we focused on generic component instruments published after the year 2000 to assess the quality of evidence [15]. A component instrument is a tool to assess all aspects which may introduce bias like randomization or type of blinding. Five instruments were used [16], [17], [18], [19], [20] which differed by the number of domains, questions within domains, and answer options within a question (see Table 1).

In a second step, the students were asked to rate the overall risk of bias on a five point Likert scale (very low, low, moderate, high, very high), based on the outcomes of the risk assessment. Each student was asked to evaluate publications of five randomized controlled trials (RCTs). Assessing five RCTs by five instruments would

lead to 25 combinations of RCTs and instruments. To reduce the efforts for the students, each RCT was assessed by one instrument only, which was randomly selected, resulting in five assessments per student. Permutation was used to ensure that each RCT was assessed with each instrument by the same number of students. Then, five experts assessed the five RCTs with a single instrument, which was selected randomly (see Table 2 for assignment of instrument and RCT to expert). These assessments were used as the reference standard ('gold standard'). Experts had to fulfill the following criteria: independence from the University of Duisburg-Essen, experience in critically appraising clinical studies for more than 10 years, and professional status including responsibility for assessments.

2.2. Training sessions

At the beginning, we presented the concept of EBM and its five steps by means of a specific clinical case and in detail. Essential terms were explained theoretically and in a traditional teaching approach: internal and external validity, quality parameters like randomization, concealment, blinding, drop-out/lost to follow-up, intention-to-treat (ITT) analysis, evidence levels of study designs, PICO-scheme and different kinds of bias. In addition, the structure of scientific publications was explained and therein text passages were indicated, where the description or discussion of validity aspects is most likely to be found. The students did a practical exercise, based on this information, afterwards. As an aid, a slide was shown containing the glossary of terms for quick searches (see Attachment 1).

The second session started with a repetition and consolidation of the knowledge gained. For this purpose, the participants were asked to split into groups and to allocate key words to quality parameters. After internal discussion, the groups used a flip-chart for poster presentation, during which the assignment of the key words to a quality aspect had to be explained to the others. For a further understanding, a simulation of randomization, blinding, concealment, stratification, drop-out/lost to follow-up and different types of analyses (e.g. ITT) was carried out thereafter.

In the third session, the component systems for the quality assessment were introduced and the single questions of the instruments were discussed. The participants then applied the systems for quasi randomly selected RCTs about the most frequent chronic diseases defined by the WHO, and results were discussed.

In the final session, the students' skills were tested by application of the component systems on another set of currently published RCT. Again, permutation was done for assigning instruments to RCTs. Analyses of the methodological study are based on these assessments.

Table 1: Component Instruments

	Higgins&Green [16]	Hill [17]	Huwiler-Müntener [18]	IQWiG [19]	Thomas ^a [20]
Domains	n=7	n=6	n=3	n=7	n=8
Randomization	Yes	Yes	No	Yes	Yes
Concealment	Yes	Yes	Yes	Yes	No
Blinding (double)	Yes ^b	Yes	Yes	No	Yes ^c
Blinding (outcome assessors)	Yes	No	No	Yes	NA
Confounding	Yes ^d	No	No	Yes	Yes
Drop-out	Yes ^e	Yes ^f	No	Yes ^g	Yes ^h
Other	Selective outcome reporting	ITT/modified ITT/Completers analysis	ITT/on treatment/unclear	Sample size calculation	Selection of participants
		Manufacturer support		Comparability of groups	Data collection methods
					Intervention integrity
					Setting/ITT

a Each domain contains an additional overall rating at the end with the options strong, moderate, weak
 Asking for:
 b Participants and personnel
 c Participants and outcome assessors
 d Other sources of bias
 e Incomplete outcome data, f drop-out, g, transparency of drop-out, h numbers and/or reasons for drop-out
 Hints to fill out available: Higgins & Green [16], Hill [17], Thomas [20]
 IQWiG: Institute for quality and efficiency in healthcare, ITT: Intention-to-treat analysis, NA: not applicable

Table 2: Assignment of Assessment Instrument and RCT to Experts

Instrument	RCT				
	Balegar [21]	Vlug [22]	Martins [27]	Mega [23]	Merenstein [34]
Higgins&Green [16]	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Hill [17]	Expert 5	Expert 1	Expert 2	Expert 3	Expert 4
Huwiler-Müntener [18]	Expert 4	Expert 5	Expert 1	Expert 2	Expert 3
IQWiG [19]	Expert 3	Expert 4	Expert 5	Expert 1	Expert 2
Thomas [20]	Expert 2	Expert 3	Expert 4	Expert 5	Expert 1

Five experts were randomly assigned to assess each RCT by one instrument using permutation
 IQWiG: Institute for quality and efficiency in healthcare

2.3. Analysis

Frequencies of assessments given by medical students and health care management students were computed by RCT and instrument. In addition, agreement between student rating and the expert's assessment was defined as "agreement +/-1": Agreement was considered as attained when the student rating was within a range of +/- 1 point of the expert's assessment. Generalized estimation equations were used to investigate the influence of the instrument, the student group and the RCT on the agreement rate in a first analysis. In a second analysis, the effect of instrument and student group was assessed by RCT. In addition, influences of "experience in critically appraising" and "command of English" were investigated. Analyses were performed with SAS 9.2 (SAS Institute Inc., Cary, NC, USA).

3. Results

In total, 167 students took part in our EBM courses, of whom 142 were third year undergraduate medical students and 25 students of the master's program "health care management" (see Table 3).

The investigation of our primary outcome, the applicability of different instruments for the assessment of methodological study quality, did not provide evidence on the comprehensibility of instruments, instructions for using the instruments, if available, or duration of assessment procedure.

With regard to our secondary outcomes, we only report the significant findings on agreement with expert assessments. Figure 1 shows the percentages of medical students, who have given the five ratings by RCT and instrument. The rating with a black bar indicates the expert rating, e.g. for RCT 1 [21] and the IQWiG instrument, the expert assessed the study as having a high potential for bias, for RCT 2 [22], the expert found a low potential for bias, when using the IQWiG instrument. The height of the bar represents the proportion of students by rating in percent. The size of the black bar indicates, how many students got the same rate as the expert. The assessments of the same study by different experts with different instruments however show some variability. To reflect this, the agreement rate was assessed by comparing the students' assessments with a tolerance of +/-1 with the expert assessments. The majority of assessments shows an agreement in a range of 66% to over 80% and therefore an adequate rate.

Table 3: Characteristics of Students

	Medical students	HCM students
Gender	n=142	n=25
male	47 (33.10%)	9 (36.00%)
female	95 (66.90%)	16 (64.00%)
Year of Birth	n=134	n=22
1978-1989	32 (23.88%)	16 (72.73%)
1990-1995	102 (76.12%)	6 (27.27%)
Command of English	n=143	n=26
Weak	10 (6.99%)	0 (0.00%)
Low	13 (9.09%)	1 (3.85%)
Intermediate	53 (37.06%)	11 (42.31%)
Good	43 (30.07%)	13 (50.00%)
High	24 (16.78%)	1 (3.85%)
Knowledge in statistics and/or epidemiology	n=142	n=26
Weak	57 (40.14%)	1 (3.85%)
Low	41 (28.87%)	9 (34.62%)
Intermediate	35 (24.65%)	13 (50.00%)
Good	5 (3.52%)	3 (11.54%)
High	4 (2.82%)	0 (0.00%)
Prior medical knowledge	Not applicable	n=26
Weak	-	1 (3.85%)
Low	-	8 (30.77%)
Intermediate	-	10 (38.46%)
Good	-	6 (23.08%)
High	-	1 (3.85%)
Experience in critically appraising	n=140	n=26
Yes	20 (14.29%)	4 (15.38%)
No	120 (85.71%)	22 (84.62%)
Experience with instruments for bias assessment	n=139	n=23
Yes	5 (3.60%)	1 (4.35%)
No	134 (96.40%)	22 (95.65%)
Knowledge gained by EBM lecture	n=140	n=26
Weak	71 (50.71%)	1 (3.85%)
Low	31 (22.14%)	2 (7.69%)
Intermediate	33 (23.57%)	10 (38.46%)
Good	1 (0.71%)	11 (42.31%)
High	4 (2.86%)	2 (7.69%)

HCM: Health care management, IQWiG: Institute for quality and efficiency in health care
n=number of students with answer given

Table 4 shows the agreement rates for “agreement +/- 1” attained by student group in the various RCTs for the five instruments under investigation. Across RCTs, evidence was found that the choice of instrument had an impact ($p=0.0158$), while no difference for an influence of student group or RCT was found ($p=0.3856$ and $p=0.2425$, respectively). By RCT, evidence was found for an influence of the instrument on the agreement rate in RCTs 1, 2 and 3 ($p=0.0146$, $p=0.0263$ and $p<0.0001$, respectively). For the endpoint “agreement +/-1”, no evidence for an influence of “experience in critically appraising” or command of English was found. To note, the description of questions showed very different levels of detail in the publications, e.g. “the allocation code was concealed in sequentially numbered, opaque, sealed envelopes” [21] versus a simple mentioning of the term “randomization” without description of methodological details [23].

4. Discussion

We found a non-systematic over- and underestimation of risk of bias compared to the experts' assessments with no clear direction in the students' assessments, what corresponds somehow to the answers given after the four short courses in EBM: 73% of the medical students rated their knowledge gained as weak or low. Our results contrast sharply with those of many other evaluations after lessons in EBM for undergraduates. In a before-after comparison, Weberschock et al. [14] observed a significant increase of performance in 124 year 3 medical undergraduates in Germany from a score of 2.37 points before the seminar and 9.85 points thereafter (99% CI [8.94; 10.77], $p<0.001$). Carrying out a controlled educational study, Ghali et al. [11] showed a significant difference in literature searching ($p<0.002$) and critical appraisal skills ($p<0.0002$) between third year medical students in Boston visiting either four sessions in EBM or receiving traditional didactic teaching in various clinical topics. A sys-

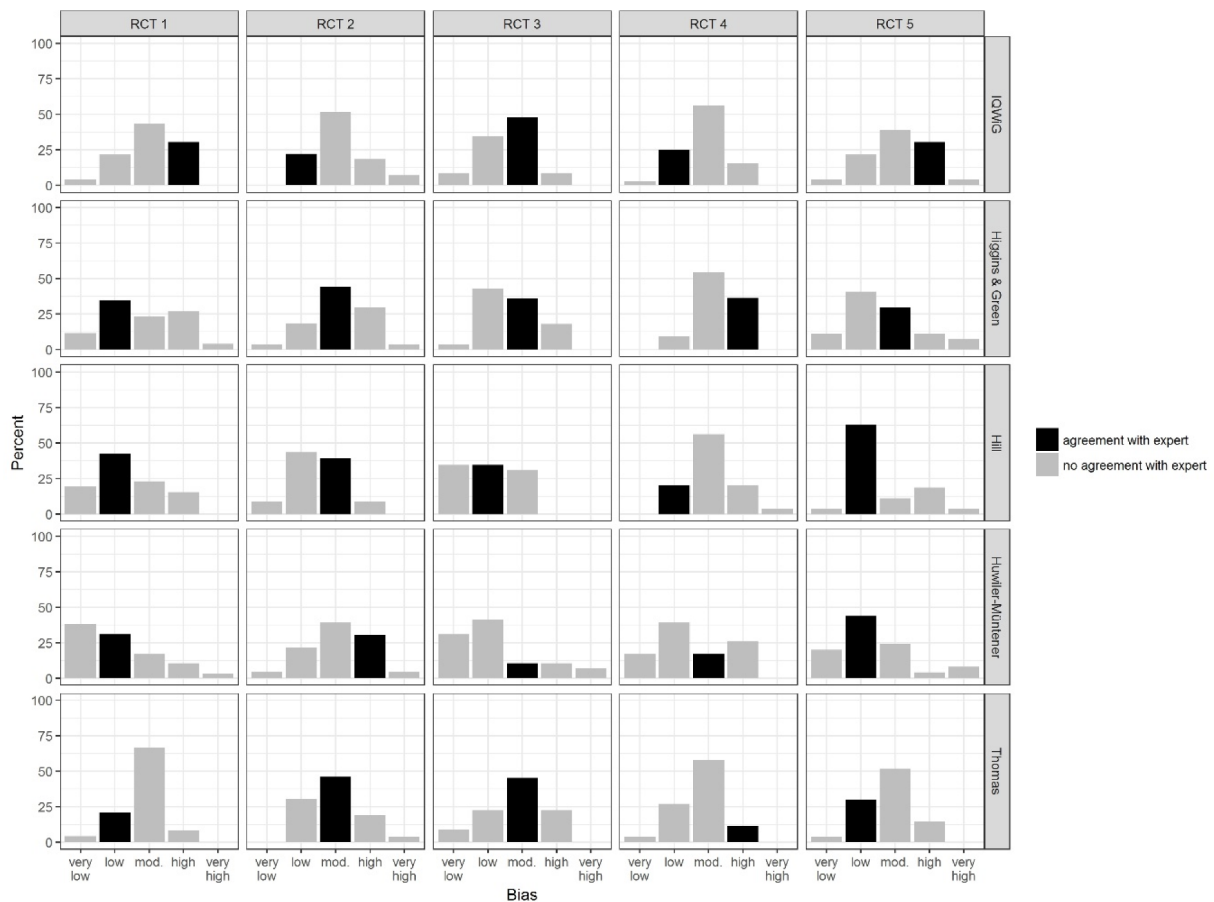


Figure 1: Assessments of medical students by RCT and instrument. Bars represent percentages. RCT 1 [21], RCT 2 [22], RCT 3 [27], RCT 4 [23], RCT 5 [34]

Table 4: Agreement between student and expert assessments

Agreement +/- 1 point on Likert scale n/N (%)			RCT				
			RCT 1 [21]	RCT 2 [22]	RCT 3 [27]	RCT 4 [23]	RCT 5 [34]
Component Instrument	Higgins & Green [16]	Medical students	18/26 (69.23%)	25/27 (92.59%)	27/28 (96.43%)	20/22 (90.91%)	22/27 (81.48%)
		HCM students	4/6 (66.67%)	4/4 (100.00%)	3/3 (100.00%)	5/5 (100.00%)	5/5 (100.00%)
	Hill [17]	Medical students	22/26 (84.62%)	21/23 (91.30%)	26/26 (100.00%)	19/25 (76.00%)	21/27 (77.78%)
		HCM students	4/5 (80.00%)	3/3 (100.00%)	5/5 (100.00%)	4/6 (66.67%)	4/4 (100.00%)
	Huwiler-Müntener [18]	Medical students	25/29 (86.21%)	17/23 (73.91%)	18/29 (62.07%)	19/23 (82.61%)	22/25 (88.00%)
		HCM students	4/5 (80.00%)	4/5 (80.00%)	2/6 (33.33%)	4/5 (80.00%)	4/4 (100.00%)
	IQWiG [19]	Medical students	17/23 (73.91%)	20/27 (74.01%)	21/23 (91.30%)	27/32 (84.38%)	17/23 (73.91%)
		HCM students	0/5 (0.00%)	5/5 (100.00%)	3/5 (60.00%)	4/4 (100.00%)	3/6 (50.00%)
	Thomas [20]	Medical students	22/24 (91.67%)	25/26 (96.15%)	20/22 (90.91%)	18/26 (69.23%)	23/27 (85.19%)
		HCM students	5/5 (100.00%)	5/5 (100.00%)	5/5 (100.00%)	2/3 (66.67%)	3/4 (75.00%)

HCM: Health care management, IQWiG: Institute for quality and efficiency in health care
 Assessment of risk of bias on Likert scale: (1) very low, (2) low, (3) moderate, (4) high, (5) very high

tematic review [4] about the impact of teaching critical appraisal skills including 10 clinical studies as well as a recently published review [9] including 14 RCTs about methods of teaching medical trainees EBM concluded an increase of learner competencies post-intervention across all studies. One reason for these differences compared to our results may be self-selection of highly

motivated participants in contrary to our students visiting compulsory courses. In addition, the objectivity of evaluations wasn't always strong, ranging between full self-perceptions, kinds of question papers with multiple-choice questions [14], and validated tests. As stated by Fritsche et al. 2002 [24] comparing the effects of EBM lectures between experts, postgraduate physicians and medical

students, an objective evaluation of courses in EBM may be difficult but essential because there is a poor correlation between subjective perception of knowledge and its objective assessment [3].

In our study, students' assessments were compared to a reference standard created by experts and therefore guaranteeing a certain objectivity, which is always prone to individual experience and individual perception; supplementary self-ratings of an increase in knowledge were recorded.

As each expert has rated each study by one instrument only, no assessment of reliability of experts was performed. It has to be mentioned here that even most of the existing assessment tools are not tested for validity and reliability [25]. With the exception of two [16], [20] this also applies for the included assessment instruments. Evaluating the Cochrane Collaboration's risk of bias tool [16], Hartling et al. [26] found a wide range of inter-rater reliability between experts on individual domains from slight to substantial (weighted $\kappa=0.13-0.74$). Working in the same institution and review team, the authors assume a much higher variability across different research units. As an explanation for the wide range of inter-rater agreement, the authors discuss the need for clear and detailed instructions to improve reliability [26].

As seen in figure 1, the different instruments led the different experts to different assessments of the same study. One reason may be the different domains, additional information asked by some instruments as well as different depth of questions. This may lead to discordant conclusions, if certain aspects are not asked for in some instruments but asked for others. To cover the uncertainty, we used a less strict definition of agreement in terms, that an assessment with +/- one point is still regarded an agreement. Selection bias in the shape of publication bias and reporting bias can be assumed for all of the publications mentioned above which reported clearly positive effects of EBM lectures.

The extent and comprehensibility to which single questions were described within publications varied widely, sometimes leading to an impossible task for our unexperienced undergraduate students and a distortion in the analyses as well. It seems striking that the description of questions published in higher ranked journals (RCT 3 and 4 [27], [23]) requiring compliance with special statements concerning reporting quality was less understandable for our students than in publications with a smaller impact. The fact that critical appraisal always includes subjectivity by interpretation as well as scoring systems which only appear to be objective due to an explicit or implicit weighting without any empirical basis [28] is worth pointing out within this context.

The number of five different assessment instruments could have been too complex. On the other hand, the repetition of terms served for a greater familiarity, and we only addressed step 3 of the EBM concept in detail, therefore focusing the knowledge transfer very strongly. In particular, the training of step 1, formulating a research question which can be operationalized, may take a sub-

stantial amount of time for undergraduate students without experience in scientific work. This also applies to step 2, the literature searches in electronic databases.

Teaching critical appraisal separately, as we did, is very common and also known as journal club, meaning that participants have to read and appraise articles critically under the guidance of an expert [4]. For the reason of keeping up to date with new evidence, clinicians have to go through many articles in every day practice, and to do this effectively, training is necessary [29]. Therefore, and as suggested as a format of teaching EBM under certain conditions [10] we focused on step three of the EBM concept. However, this evokes other difficulties as critical appraisal integrates knowledge about epidemiology, information science and biostatistics [9].

Although lectures in epidemiology and psychology including statistics are compulsory in the first two years in Germany, 69% of our medical students reported a weak or low knowledge in statistics and/or epidemiology, showing that attitude and knowledge are not spread in the same manner, and teaching EBM must address the needs of different learners [8]. Maybe, using step 1 and 2 of the concept as an introduction is more appropriate to foster a scientific mindset. Alternatively and in order to escape the charge of isolation from clinical practice [30], teaching step 4, applying the evidence on individual cases, may be considered. However, this cannot succeed in undergraduates when practical experience is lacking, and the concrete objective only a vague idea. This is underlined by long-term experiences from Duke and Stanford resulting in a curriculum with the precondition of a clinical training prior to research experience because students were better prepared to understand the clinical and translational potential of their research projects [31]. For an open mind and a better assessment it would also be beneficial if students were familiar with the whole development process of a clinical trial and the impact of its single aspects [32].

In order to get and apply the best available evidence to clinical decision making, skills in finding and critically appraising medical literature are an essential prerequisite [13]. Without background knowledge in methodology and statistics, physicians are at a high risk of misinterpreting evidence, leading to medical errors and adverse effects [28].

Swift et al. [6] investigated the views of 130 physicians about training in statistics and its need in daily practice. As a student, more than half of the participants (60%) underestimated the value of these subjects as relevant to medical practice whilst the majority (73%) had realized its impact for their career over time. Despite the increasing conviction in the relevance of EBM there is evidence for continuing knowledge gaps in basic statistical concepts among practicing physicians and medical researchers [32]. Likewise, a sound knowledge of key methodological EBM terms and sources seems to be lacking among the majority of health personal including physicians, translational researchers, nurses and other health professionals [12], [13]. To remedy this situation, even

students must be helped to perceive these subjects as important to clinical practice [32]. It's hoped that future physicians will better appraise research findings and contribute to furthering the clinical field by conducting research [33].

Acknowledgement

We thank Tobias Goeke, Angelika Gohlke, Anja Hagen, Beate Lux, and Monika Nothacker for their expert assessments.

Ethical approval

The seminar content and structure was approved by the office of the Dean. There was no contact with patients.

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from

<http://www.egms.de/en/journals/zma/2018-35/zma001171.shtml>

- Attachment_1.pdf (56 KB)
Glossary of terms for quick searches

References

- Godlee F. Milestones on the long road to knowledge. *BMJ*. 2007;344:s2. DOI: 10.1136/bmj.39062.570856.94
- Meskó B. Medical Milestones on the long road to knowledge. [Internet]. 2007. [cited 2015 Sept 09]. Zugänglich unter/available from <http://sciencereoll.com/2007/01/18/medical-milestones-on-the-long-road-to-knowledge/>
- Khan KS, Awonuga A, Dwarakanath LS, Taylor R. Assessments in evidence-based medicine workshops: loose connection between perception of knowledge and its objective assessment. *Med Teach*. 2001;23(1):92-94. DOI: 10.1080/01421590150214654
- Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *Can Med Assoc J*. 1998;158(2):177-181.
- Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart T. Teaching critical appraisal skills in healthcare settings. *Cochrane Database Syst Rev*. 2011;(11):CD001270. DOI: 10.1002/14651858.CD001270pub2
- Swift L, Miles S, Price GM, Shepstone L, Leinster SJ. Do doctors need statistics? Doctors' use of and attitudes to probability and statistics. *Stat Med*. 2009;28(15):1969-1981. DOI: 10.1002/sim.3608
- Altman DG, Bland JM. Improving doctors' understanding of statistics. *J Royal Stat Soc*. 1991;154(2):223-267. DOI: 10.2307/2983040
- Straus SE, Green ML, Bell DS, Badgett R, Davis D, Gerrity M, Ortiz E, Shaneyfeldt TM, Whelan C, Mangrulkar R; Society of General Internal Medicine Evidence-Based Medicine Task Force. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ*. 2004;329(7473):1029-1032. DOI: 10.1136/bmj.329.7473.1029
- Ilic D, Maloney S. Methods of teaching medical trainees evidence-based medicine: a systematic review. *Med Educ*. 2014;48(2):124-135. DOI: 10.1111/medu.12288
- Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicines changes anything? A systematic review. *BMJ*. 2004;329(7473):1017. DOI: 10.1136/bmj.329.7473.1017
- Ghali WA, Saitz R, Eskew AH, Gupta M, Quan H, Hershman WY. Successful teaching in evidence-based medicine. *Med Educ*. 2000;34(1):18-22. DOI: 10.1046/j.1365-2923.2000.00402.x
- Young JM, Glasziou P, Ward JE. General practitioners' self rating of skills in evidence based medicine: validation study. *BMJ*. 2002;324(7343):950-951. DOI: 10.1136/bmj.324.7343.950
- Ugolini D, Casanova G, Ceppi M, Mattei F, Neri M. Familiarity of physicians, translational researchers, nurses, and other health professionals with evidence-based medicines terms and resources. *J Canc Educ*. 2014;29(3):514-521. DOI: 10.1007/s13187-014-0631-0
- Weberschock TB, Ginn TC, Reinhold J, Strametz R, Krug D, Bergold M, Schulze J. Change in knowledge and skills of Year 3 undergraduates in evidence-based medicine seminars. *Med Educ*. 2005;39(7):665-671. DOI: 10.1111/j.1365-2929.2005.02191.x
- Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Schriftenreihe Health Technology Assessment, Bd. 102. Köln: Deutsches Institut für Medizinische Dokumentation und Information (DIMDI); 2010.
- Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* 5.1.0. Oxford, UK: The Cochrane Collaboration; 2011.
- Hill CL, La Valley MP, Felson DT. Secular changes in the quality of published randomized clinical trials in Rheumatology. *Arthritis Rheumatism*. 2002;46(3):779-784. DOI: 10.1002/art.512
- Huwiler-Müntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002;287(21):2801-2804. DOI: 10.1001/jama.287.21.2801
- IQWiG. Früherkennungsuntersuchung von Sehstörungen bei Kindern bis zur Vollendung des 6. Lebensjahres. Abschlussbericht. Nr. 32. Köln: IQWiG-Berichte; 2008.
- Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldview Evidence-based Nurs*. 2004;1(3):176-184. DOI: 10.1111/j.1524-475X.2004.04006.x
- Balegar VK, Kluckow M. Furosemide for packed red cell transfusion in preterm infants: a randomized controlled trial. *J Pediatr*. 2011;59(6):913-918. DOI: 10.1016/j.jpeds.2011.05.022
- Vlug MS, Wind J, Hollmann MW, Ubbink DT, Cense HA, Engel A, Gerhards MF, van Wagenveld BA, van der Zaag ES, van Geloven AA, Sprangers MA, Cuesta MA, Bemelman WA, LAFA study group. Laparoscopy in combination with fast track multimodal management is the best perioperative strategy in patients undergoing colonic surgery. *Ann Surg*. 2011;254(6):868-875. DOI: 10.1097/SLA.0b013e31821fd1ce

23. Mega JL, Braunwald E, Wiviott SD, Bassand JP, Bhatt DL, Bode C, Burton P, Cohen M, Cook-Bruno M, Fox KAA, Goto S, Murphy SA, Plotnikov AN, Schneider D, Sun X, Verheugt FW, Gibson CM. ATLAS ACS 2-TIMI 51 Investigators. *N Engl J Med*. 2012;366(1):9-19. DOI: 10.1056/NEJMoa1112277
24. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ*. 2002;325:1338-1341. DOI: 10.1136/bmj.325.7376.1338
25. Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. Validity and inter-rater reliability testing of quality assessment instruments. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021-I.) AHRQ Publication No. 12-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012. Zugänglich unter/available from: www.effectivehealthcare.ahrq.gov/reports/final.cfm
26. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339:b4012. DOI: 10.1136/bmj.b4012
27. Martins N, Morris, Kelly PM. Food incentives to improve completion of tuberculosis treatment: randomized trial in DILI, Timor-Leste. *BMJ*. 2009;339:b4248. DOI: 10.1136/bmj.b4248
28. Buchberger B, von Elm E, Gartlehner G, Huppertz H, Antes G, Wasem J, Meerpohl JJ. Assessment of risk of bias in controlled studies. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz*. 2014;57(12):1432-1438. DOI: 10.1007/s00103-014-2065-6
29. Kulier R, Gee H, Khan K. Five steps from evidence to effect: exercising clinical freedom to implement research findings. *BJOG*. 2008;115:1197-1202. DOI: 10.1111/j.1471-0528.2008.01821.x
30. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, Germar MJ, Pisake L, Mittal S, Pattinson R, Wolomy-Molondo JJ, Bergh AM, May W, Souza JP, Koppenhoefer S, Khan KS. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training. *JAMA*. 2012;308(21):2218-2225. DOI: 10.1001/jama.2012.33640
31. Laskowitz DT, Drucker RP, Parsonnet J, Cross PC, Gesundheit N. Engaging students in dedicated research and scholarship during medical school: the long term experiences at Duke and Stanford. *Acad Med*. 2010;85(3):419-428. DOI: 10.1097/ACM.0b013e3181ccc77a
32. Miles S, Price GM, Swift L, Shepstone L, Leinster SJ. Statistics teaching in medical school: opinion of practicing doctors. *BMC Med Educ*. 2010;10:75. DOI: 10.1186/1472-6920-10-75
33. Vereijken MW, Kruidering-Hall M, de Jong PG, de Beaufort AJ, Dekker FW. Scientific education early in the curriculum using a constructivist approach on learning. *Perspect Med Educ*. 2013;2(4):209-215. DOI: 10.1007/s40037-013-0072-1
34. Merenstein D, Murphy M, Fokar A, Hernandez RK, Park H, Nsouli H, Sanders ME, Davis BA, Niborski V, Tondu F, Shara NM. Use of fermented dairy probiotic drink containing lactobacillus casei (DN-114 001) to decrease the rate of illness in kids: the DRINK study. A patient-oriented, double-blind, cluster-randomized, placebo-controlled, clinical trial. *Eur J Clin Nutr*. 2010;64(7):669-677. DOI: 10.1038/ejcn.2010.65

Corresponding author:

B. Buchberger

University of Duisburg-Essen, Faculty of Economics and Business Administration, Institute for Health Care Management and Research, Thea-Leymann-Str. 9, D-45127 Essen, Germany, Phone: +49 (0)201/183-4075, Fax: +49 (0)201/183-4073
barbara.buchberger@medma.uni-due.de

Please cite as

Buchberger B, Mattivi JT, Schwenke C, Katzer C, Huppertz H, Wasem J. Critical appraisal of RCTs by 3rd year undergraduates after short courses in EBM compared to expert appraisal. *GMS J Med Educ*. 2018;35(2):Doc24. DOI: 10.3205/zma001171, URN: urn:nbn:de:0183-zma0011712

This article is freely available from

<http://www.egms.de/en/journals/zma/2018-35/zma001171.shtml>

Received: 2017-03-21**Revised:** 2017-11-25**Accepted:** 2018-01-31**Published:** 2018-05-15**Copyright**

©2018 Buchberger et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Kritische Bewertung von RCTs nach einem EBM-Grundkurs im dritten Studienjahr im Vergleich zur Expertenbewertung

Zusammenfassung

Einleitung: Ein wesentliches Ziel von Kursen in evidenzbasierter Medizin (EBM) ist die Verbesserung von Fähigkeiten für adäquates Lesen und Interpretieren medizinischer Fachliteratur. Hinsichtlich der Konzeption ist es wichtig, unterschiedliche Ausbildungsgrade zu berücksichtigen.

Ziel: Primäres Ziel war die Untersuchung der Anwendbarkeit verschiedener Instrumente zur Bewertung der methodischen Qualität von publizierten Studien durch Studenten im dritten Studienjahr nach einem Kurzlehrgang in EBM. Sekundäre Ziele waren Übereinstimmung mit Experten-Bewertungen sowie Kenntnisse und Kompetenzen der Studenten.

Methoden: Vier Veranstaltungen zur EBM von jeweils 90 Minuten wurden für Medizinmanagement- und Medizinstudenten mit dem Fokus auf die kritische Bewertung von Literatur in Hinblick auf ein Bias-Risiko durchgeführt. Am Ende bewerteten die Studenten fünf Publikationen über randomisierte klinische Studien (RCTs) mit fünf verschiedenen Instrumenten; die Ergebnisse wurden mit Bewertungen von Experten verglichen.

Ergebnisse: Insgesamt nahmen 167 Studenten an den EBM-Kursen teil. Im Vergleich zu den Experten-Bewertungen zeigten die Bewertungen der Studenten eine nicht systematische Über- und Unterschätzung des Bias-Risikos ohne erkennbare Richtung. Die Übereinstimmung mit den Experten-Bewertungen reichte von 66% bis über 80%. Für die RCTs zeigte sich, dass die Auswahl des Instruments einen Einfluss auf die Übereinstimmungsrate zwischen studentischen und Experten-Bewertungen hatte ($p=0,0158$). Bei drei RCTs konnte ein Einfluss des Instruments auf die Übereinstimmungsrate festgestellt werden (jeweils $p<0,05$).

Diskussion: Die Ergebnisse stehen in starkem Widerspruch zu denen vieler anderer, vergleichbarer Erhebungen. Gründe können in der mangelnden Motivation der Studenten für die Pflichtveranstaltung liegen. Es ist aber auch möglich, dass der Vergleich mit einem Referenzstandard als Ergänzung zu Selbstauskünften zu Objektivität führte.

Schlussfolgerung: Nicht graduierte Studenten sollten mit den Prinzipien der EBM so früh wie möglich vertraut gemacht werden, ebenso mit Forschungsmethoden und der Lektüre von wissenschaftlichen Publikationen. Für ein tieferes Verständnis scheint klinische Erfahrung eine unverzichtbare Voraussetzung zu sein. Auf der Basis unserer Ergebnisse empfehlen wir, Veranstaltungen zur EBM und kritischen Bewertung mindestens zweimal im Studienverlauf und mit größerer Intensität kurz vor dem Abschluss durchzuführen.

Schlüsselwörter: Kritische Bewertung, evidenzbasierte Medizin, randomisierte kontrollierte Studie, Training, nicht graduierte Studenten

1. Einleitung

Der Bekanntheitsgrad von evidenzbasierter Medizin (EBM) nimmt weltweit zu, und die Akzeptanz für das Konzept wächst. Im Januar 2007 führte das British Medical Journal

eine Online-Abstimmung zu den 15 bedeutendsten medizinischen Meilensteinen durch, und EBM erreichte Platz sieben, unmittelbar hinter Keimtheorie und oraler Empfängnisverhütung, aber vor Computern und medizinischer Bildgebung [1], [2]. Die Förderung einer Kultur der EBM und ihre Implementierung in die Praxis setzt jedoch die

B. Buchberger¹

J.T. Mattivi¹

C. Schwenke²

C. Katzer¹

H. Huppertz¹

J. Wasem¹

1 Universität Duisburg-Essen, Fakultät für Wirtschaftswissenschaften, Lehrstuhl für Medizinmanagement, Essen, Deutschland

2 SCO:SSiS, Schwenke Consulting: Strategies and Solutions in Statistics, Berlin, Deutschland

Fähigkeit zur Identifizierung und kritischen Bewertung von Literatur voraus [3], [4], [5]. Gewisse Kenntnisse über Wahrscheinlichkeit und Statistik sind dafür zwingend erforderlich, wie auch für das Verständnis von Leitlinien und Evidenz-Synthesen, die Bewertung von Marketing- und Werbemaßnahmen der Industrie, Interpretation von Ergebnissen eines Screening-Tests oder der Lektüre von wissenschaftlichen Publikationen, um auf dem neuesten Stand der Entwicklung von Behandlungen zu bleiben; darüber hinaus sind Kenntnisse der Biostatistik notwendig, um numerische Daten zu analysieren, Patienten über Behandlungsrisiken aufzuklären, und nicht zuletzt, um auf Situationen vorbereitet zu sein, in denen Patienten Literatur unterschiedlichster Qualität aus dem Internet präsentieren [6], [7]. Tatsächlich stellt sich die Frage nicht länger, ob EBM unterrichtet werden soll, sondern wie [8] und wann. Abgesehen von verschiedenen pädagogischen Methoden wie z.B. „Training-on-the-job“, problembasiertem oder selbstgesteuertem Lernen [9] kann das Konzept der EBM als Ganzes, aber auch getrennt nach einzelnen der fünf Schritte vermittelt werden; diese sind:

1. Stellen einer klinischen Frage,
2. Suche nach der besten Evidenz,
3. kritische Bewertung der Evidenz,
4. Anwenden der Evidenz bei Patienten,
5. Selbstbewertung [8].

Es gibt eine große Anzahl von Kursen zur Einführung in Datenbankrecherchen durch Bibliothekare [9] oder in Lerngruppen zum Trainieren der kritischen Bewertung von Literatur [9], [10], [11]. Um den Wissenszuwachs durch den Besuch von EBM-Kursen zu messen, sind objektive Methoden von größerem Wert als Selbstbewertungen, die zu erheblichen Überschätzungen führen [12], [13].

In früheren Studien wurde der Effekt von EBM-Kursen überwiegend durch Selbstauskünfte der Teilnehmer gemessen [4], [11] oder in Form von Fragebögen mit Multiple-Choice-Fragen [14]. Zusätzlich zu Selbstbewertungen war das Ziel der vorliegenden Studie eine gewisse Objektivität infolge des Vergleichs von Bewertungen durch Studenten mit Experten-Bewertungen als Referenzstandard.

Primäres Ziel war die Untersuchung der Anwendbarkeit verschiedener Instrumente zur Bewertung der methodischen Qualität von publizierten Studien durch Studenten im dritten Studienjahr nach einem Kurzlehrgang in EBM. Sekundäre Ziele waren Übereinstimmung mit Experten-Bewertungen sowie Kenntnisse und Kompetenzen der Studenten.

2. Methoden

Es wurden Medizinstudenten der medizinischen Fakultät unmittelbar nach Abschluss der Vorklinik und Studenten im Masterprogramm Medizinmanagement der wirtschaftswissenschaftlichen Fakultät der Universität Duisburg-Essen in die Untersuchung eingeschlossen. Im Kontext einer

Vorlesungsreihe zur Gesundheitsökonomie, durchgeführt durch den Lehrstuhl für Medizinmanagement und Pflichtveranstaltung für Medizinstudenten, wurden den Studenten im Wintersemester 2013/2014 in vier Veranstaltungen zu jeweils 90 Minuten die Prinzipien der EBM vermittelt. Die Anzahl der Teilnehmer war auf 20 Studenten pro Termin beschränkt. Ein Glossar mit Begriffen zur Schnellsuche wurde ausgehändigt (siehe Anhang 1).

2.1. Beschreibung von Instrumenten und Experten

Auf der Basis eines Health Technology Assessment Berichts wurde der Fokus auf generische Instrumente publiziert nach dem Jahr 2000 zur Komponentenbewertung der Qualität von Evidenz gerichtet [15]. Mit einem Instrument zur Komponentenbewertung werden alle Aspekte, die zu Bias führen können, wie z.B. Randomisierung oder Art der Verblindung erfasst. Fünf Instrumente wurden eingesetzt [16], [17], [18], [19], [20], die sich in der Anzahl von Domänen, Fragen innerhalb von Domänen und Antwortmöglichkeiten unterscheiden (siehe Tabelle 1). In einem zweiten Schritt wurden die Studenten gebeten, das Gesamtrisiko für Bias auf einer fünfstufigen Likert-Skala (sehr gering, gering, mäßig, hoch, sehr hoch) auf der Basis der einzelnen Biasrisiko-Bewertungen zu beurteilen. Jeder Student hatte fünf Publikationen über fünf randomisierte kontrollierte Studien (RCTs) auszuwerten. Die Bewertung von fünf RCTs durch fünf Instrumente ergab 25 Kombinationen von RCTs mit Instrumenten. Um den Aufwand für die Studenten zu reduzieren, musste jede RCT nur mit einem Instrument, das zufällig ausgewählt war, bewertet werden, so dass jeder Student lediglich fünf Auswertungen vornehmen musste. Dabei wurde mithilfe einer Permutation sichergestellt, dass jede RCT mit jedem Instrument durch dieselbe Anzahl von Studenten bewertet wurde.

Parallel dazu wurden die fünf RCTs von fünf Experten mit jeweils einem Instrument, das zufällig zugeteilt wurde, bewertet (siehe Tabelle 2 zur Zuordnung von Bewertungsinstrument und RCT zu Experten). Diese Bewertungen dienten als Referenz-Standard (Goldstandard). Die Experten mussten folgende Kriterien erfüllen: Unabhängigkeit von der Universität Duisburg-Essen, mehr als zehnjährige Erfahrung in der kritischen Bewertung von klinischen Studien, berufliche Position mit der Verantwortung für derartige Bewertungen.

2.2. Veranstaltungen

Zu Beginn wurde das Konzept der EBM und seine fünf Schritte anhand eines Einzelfallbeispiels detailliert vorgestellt. Wesentliche Begriffe wurden theoretisch und nach traditionellem pädagogischen Ansatz erklärt: interne und externe Validität, Qualitätsparameter wie Randomisierung, verdeckte Gruppenzuteilung, Verblindung, Drop-out/lost to follow-up, intention-to-treat (ITT)-Analyse, Evidenzlevel von Studiendesigns, PICO-Schema und verschiedene Arten von Bias. Zusätzlich wurde die Struktur von wissen-

Tabelle 1: Instrumente zur Komponentenbewertung

	Higgins&Green [16]	Hill [17]	Huwiler-Müntener [18]	IQWiG [19]	Thomas ^a [20]
Domäne	n=7	n=6	n=3	n=7	n=8
Randomisierung	Ja	Ja	Nein	Ja	Ja
Verdeckte Gruppenzuteilung (Concealment)	Ja	Ja	Ja	Ja	Nein
Verblindung (doppelt)	Ja ^b	Ja	Ja	Nein	Ja ^c
Verblindung (Endpunkterhebung)	Ja	Nein	Nein	Ja	NA
Confounding	Ja ^d	Nein	Nein	Ja	Ja
Drop-out	Ja ^e	Ja ^f	Nein	Ja ^g	Ja ^h
Andere	Selektives Berichten zu Endpunkten	ITT/modifizierte ITT/ Analyse vollständiger Protokoll-erfüllung	ITT/nach Behandlung/ unklar	Fallzahlplanung	Auswahl von Teilnehmern
		Unterstützung durch Hersteller		Vergleichbarkeit der Gruppen	Methode der Datenerhebung
					Vollständigkeit der Intervention
					Setting/ITT

a Eine Gesamtbewertung jeder Domäne erfolgt jeweils am Ende mit: stark, moderat oder schwach
 Frage nach:
 b Teilnehmer und Studienpersonal
 c Teilnehmer und Endpunkterhebung
 d Andere Ursachen für Bias
 e Unvollständige Daten, f Drop-out, g Transparenz Drop-out, h Angabe Anzahl und/oder Gründe für Drop-out
 Ausfüllhinweise gegeben: Higgins & Green [16], Hill [17], Thomas [20]
 IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, ITT: Intention-to-treat-Analyse, NA: nicht anwendbar

Tabelle 2: Zuordnung von Bewertungsinstrument und RCT zu Experten

Instrument	RCT				
	Balegar [21]	Vlug [22]	Martins [27]	Mega [23]	Merenstein [34]
Higgins&Green [16]	Experte 1	Experte 2	Experte 3	Experte 4	Experte 5
Hill [17]	Experte 5	Experte 1	Experte 2	Experte 3	Experte 4
Huwiler-Müntener [18]	Experte 4	Experte 5	Experte 1	Experte 2	Experte 3
IQWiG [19]	Experte 3	Experte 4	Experte 5	Experte 1	Experte 2
Thomas [20]	Experte 2	Experte 3	Experte 4	Experte 5	Experte 1

Fünf Experten wurden mittels Permutation randomisiert zugeordnet, um jede RCT durch jeweils ein Instrument zu bewerten
 IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

schaftlichen Publikationen erläutert und auf Textabschnitte hingewiesen, in denen die Beschreibung oder Diskussion von Validitätsaspekten üblicherweise erfolgt.

Im Anschluss daran führten die Studenten eine praktische Übung auf Basis der gegebenen Informationen durch. Als Hilfe wurde eine Folie gezeigt, auf der ein Glossar mit Begriffen zur Schnellsuche abgebildet war (siehe Anhang 1).

Die zweite Veranstaltung begann mit einer Wiederholung und Vertiefung der gewonnenen Kenntnisse. Zu diesem Zweck wurden die Teilnehmer gebeten, sich in Gruppen aufzuteilen und Schlüsselbegriffe Qualitätsparametern zuzuordnen. Im Anschluss an interne Diskussionen fand mittels Flipchart eine Posterpräsentation statt, während der die zugeordneten Begriffe und Qualitätsparameter den jeweils anderen Gruppen erläutert werden mussten. Zum weiteren Verständnis wurden Simulationen von Randomisierung, Verblindung, verdeckter Gruppenzuteilung, Stratifizierung, Drop-out/ lost to follow-up und verschiedenen Arten von Analysen wie z.B. ITT durchgeführt. In der dritten Veranstaltung wurden zunächst die Komponentensysteme zur Qualitätsbewertung vorgestellt und

einzelne Fragen zu den Instrumenten diskutiert. Anschließend wendeten die Teilnehmer die Systeme auf quasi zufällig ausgewählte RCTs zu den häufigsten chronischen Erkrankungen gemäß Definition der WHO an, und die Ergebnisse wurden diskutiert.

In der letzten Veranstaltung wurden die Fähigkeiten der Studenten durch Anwendung der Komponentensysteme auf ein weiteres Set von kürzlich publizierten RCTs getestet. Auch hier wurde mittels Permutation sichergestellt, dass die Zuordnung von Instrumenten zu den RCTs balanciert war. Die Analysen der methodischen Studie basieren auf diesen Bewertungen.

2.3. Analyse

Die Häufigkeiten der Bewertungen der Medizin- und Medizinmanagementstudenten wurden für jede RCT und jedes Instrument berechnet. Zusätzlich wurden die Übereinstimmungsraten zwischen den Teilnehmern und den Experten bestimmt. Eine Übereinstimmung lag vor, wenn die Bewertung des Teilnehmers um maximal einen Punkt auf der Likert-Skala von der Bewertung des Exper-

ten abwich. Generalisierte Schätzgleichungen (Generalized estimation equations) wurden verwendet, um den Einfluss des Instruments, der Studentengruppe und der RCT auf die Übereinstimmungsrate zu untersuchen. In einer zweiten Analyse wurde der Effekt des Instruments und der Studentengruppe pro RCT untersucht. In einer weiteren Analyse wurde der Einfluss der Faktoren „Erfahrung mit kritischen Bewertungen“ und „Englischkenntnisse“ untersucht. Alle Analysen wurden mit der Software SAS 9.2 (SAS Institute Inc., Cary, NC, USA) durchgeführt.

3. Ergebnisse

Insgesamt nahmen 167 Studenten an den EBM-Kursen teil, von denen 142 Medizinstudenten im dritten Studienjahr und 25 Studenten des Masterprogramms Medizinmanagement waren (siehe Tabelle 3).

Die Untersuchung des primären Zielparameters, Anwendbarkeit verschiedener Instrumente zur Bewertung von methodischer Studienqualität, erbrachte weder Evidenz zur Verständlichkeit der Instrumente, noch zu den ggf. vorhandenen Hinweisen zum Gebrauch der Instrumente oder zur Dauer des Bewertungsprozesses.

Hinsichtlich der sekundären Zielparameter werden lediglich signifikante Ergebnisse zur Übereinstimmung mit Experten-Bewertungen berichtet. In Abbildung 1 sind die prozentualen Anteile der fünf Bewertungen nach RCT und Instrument der Medizinstudenten dargestellt.

Der schwarze Balken stellt die Expertenbewertung dar; z.B. identifizierte der Experte für RCT 1 [21] ein hohes Potential für Bias, für RCT 2 [22] mit dem IQWiG-Instrument ein niedriges Potential für Bias. Die Höhe des Balkens zeigt den Anteil der Bewertungen der Studenten gemäß der Likert-Skala. Die Höhe des schwarzen Balkens zeigt, wieviele Studenten die gleiche Bewertung wie der Experte wählten. Die Bewertungen der gleichen Studie durch verschiedene Experten mit verschiedenen Instrumenten wiesen einige Variabilität auf. Um dem Rechnung zu tragen, wurde als Übereinstimmung der Studentenbewertungen mit den Expertenbewertungen eine Toleranz von +/-1 Punkt auf der Likert-Skala als gleich akzeptiert. Der Großteil der Bewertungen zeigte eine Übereinstimmungsrate im Bereich von 66% bis über 80% und wurde als adäquat angesehen.

In Tabelle 4 sind die Übereinstimmungsraten für die „Übereinstimmung mit +/-1“ nach Studentengruppe für die verschiedenen RCTs und die fünf Instrumente dargestellt. Über die RCTs hinweg zeigte sich ein deutlicher Effekt durch die Wahl des Instruments auf die Übereinstimmung ($p=0.0158$), wohingegen die Studentengruppe und die verschiedenen RCTs keinen Einfluss zeigten ($p=0.3856$, bzw. $p=0.2425$). In den Analysen nach RCT zeigte sich ein Einfluss des Instruments auf die Übereinstimmungsrate in den RCTs 1, 2 und 3 ($p=0.0146$, $p=0.0263$ und $p<0.0001$). Für den Endpunkt „Übereinstimmung +/-1“ wurde kein Einfluss der Faktoren „Erfahrung mit kritischen Bewertungen“ und „Englischkenntnis-

se“ gefunden. Zu beachten ist dabei, dass viele Details in den Publikationen mit sehr unterschiedlicher Detailtiefe beschrieben waren. So wurde z.B. zur Randomisierung in einem Fall ausgeführt: „the allocation code was concealed in sequentially numbered, opaque, sealed envelopes“ [21], in einer anderen Publikation wurde lediglich erwähnt, dass eine Randomisierung durchgeführt wurde ohne Nennung weiterer Details [23].

4. Diskussion

Eine nicht systematische Über- und Unterschätzung des Bias-Risikos ohne erkennbare Richtung konnte im Vergleich der Experten-Bewertungen mit Studenten-Bewertungen festgestellt werden. Dies entspricht auf gewisse Weise den Antworten, die die Studenten am Ende der vier Veranstaltungen zur EBM auf die Frage nach gewonnenen Kenntnissen gaben: 73% der Medizinstudenten bewerteten diese als schwach oder gering. Unsere Ergebnisse stehen in starkem Widerspruch zu denen vieler anderer Erhebungen am Ende von EBM-Kursen für nicht graduierte Studenten. In einem Vorher-Nachher-Vergleich konnten Weberschock et al. [14] einen signifikanten Leistungsanstieg bei 124 deutschen Medizinstudenten im dritten Studienjahr von 2,37 Score-Punkten vor dem Seminar und 9,85 Score-Punkten nach dem Seminar beobachten (95% KI [8,94; 10,77], $p<0,001$). In einer kontrollierten Studie zeigten Ghali et al. [11] einen signifikanten Unterschied für Literaturrecherche ($p<0,002$) und Fähigkeiten zur kritischen Bewertung ($p<0,0002$) zwischen Medizinstudenten aus Boston, die entweder vier Veranstaltungen zur EBM oder traditionelle Lehrveranstaltungen zu unterschiedlichen klinischen Themengebieten besucht hatten. Einem systematischen Review [4] über den Effekt von Lehrveranstaltungen zu kritischer Bewertung mit zehn eingeschlossenen klinischen Studien zufolge sowie einem vor kurzem publizierten Review [9] mit 14 eingeschlossenen RCTs über Lehrmethoden der EBM für Medizinstudenten konnte in allen Studien ein Anstieg der Kompetenzen nach der Intervention festgestellt werden. Ein Grund für diese Unterschiede im Vergleich zu den Ergebnissen der vorliegenden Studie kann die Selbstselektion hoch motivierter Teilnehmer im Gegensatz zu unseren Studenten, die eine Pflichtveranstaltung besuchten, sein. Zusätzlich dazu war die Objektivität der Evaluationen nicht immer gegeben, die aus vollständigen Selbstauskünften, einer Art Fragebogen mit Multiple-Choice-Fragen [14] und validierten Tests bestanden. Wie von Fritsche et al. 2002 [24] angeführt, die die Effekte von EBM-Kursen auf Experten, postgraduierte Ärzte und Medizinstudenten verglichen, ist eine objektive Evaluation von EBM-Kursen schwierig aber unentbehrlich, denn die Korrelation zwischen subjektiver Wahrnehmung von Wissen und objektiver Bewertung ist schwach [3]. In unserer Studie wurden die Bewertungen der Studenten mit einem Referenzstandard verglichen, der von Experten gesetzt war und daher eine gewisse Objektivität garantierte, die zugegebenermaßen immer auch durch individuelle

Tabelle 3: Charakteristika der Studenten

	Medizinstudenten	MedMan-Studenten
Geschlecht	n=142	n=25
männlich	47 (33.10%)	9 (36.00%)
weiblich	95 (66.90%)	16 (64.00%)
Geburtsjahr	n=134	n=22
1978-1989	32 (23.88%)	16 (72.73%)
1990-1995	102 (76.12%)	6 (27.27%)
Englischkenntnisse	n=143	n=26
Schwach	10 (6.99%)	0 (0.00%)
Gering	13 (9.09%)	1 (3.85%)
Mittelmäßig	53 (37.06%)	11 (42.31%)
Gut	43 (30.07%)	13 (50.00%)
Sehr gut	24 (16.78%)	1 (3.85%)
Statistik- und/oder Epidemiologiekennntnisse	n=142	n=26
Schwach	57 (40.14%)	1 (3.85%)
Gering	41 (28.87%)	9 (34.62%)
Mittelmäßig	35 (24.65%)	13 (50.00%)
Gut	5 (3.52%)	3 (11.54%)
Sehr gut	4 (2.82%)	0 (0.00%)
Medizinische Vorkenntnisse	Nicht anwendbar	n=26
Schwach	-	1 (3.85%)
Gering	-	8 (30.77%)
Mittelmäßig	-	10 (38.46%)
Gut	-	6 (23.08%)
Sehr gut	-	1 (3.85%)
Erfahrung in kritischer Bewertung	n=140	n=26
Ja	20 (14.29%)	4 (15.38%)
Nein	120 (85.71%)	22 (84.62%)
Erfahrung mit Bias-Bewertung	n=139	n=23
Ja	5 (3.60%)	1 (4.35%)
Nein	134 (96.40%)	22 (95.65%)
Gewonnene Kenntnisse durch EBM-Kurs	n=140	n=26
Schwach	71 (50.71%)	1 (3.85%)
Gering	31 (22.14%)	2 (7.69%)
Mittelmäßig	33 (23.57%)	10 (38.46%)
Gut	1 (0.71%)	11 (42.31%)
Sehr gut	4 (2.86%)	2 (7.69%)
MedMan: Medizinmanagement, IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen n=Anzahl von Studenten mit Antwort		

Erfahrungen und Ansichten geprägt ist; zusätzlich wurden Eigenangaben über den Wissenszuwachs erhoben. Da jeder Experte nur jeweils eine Studie mit einem Instrument bewertete, konnte die Reliabilität nicht getestet werden. An dieser Stelle sei erwähnt, dass Reliabilität und Validität der meisten gängigen Bewertungsinstrumente nicht überprüft sind [25]. Mit Ausnahme von zwei Instrumenten [16], [20] trifft das auch auf die in der vorliegenden Studie verwendeten Instrumente zu. Bei der Bewertung des Risk-of-Bias-Tool der Cochrane Collaboration [16] stellten Hartling et al. [26] eine große Spannweite der Interrater-Reliabilität bei Expertenbewertungen in einzelnen Domänen fest, die von leicht bis substantiell reichte (gewichteter $\kappa=0,13-0,74$). Aufgrund der Tatsache, dass die Autoren alle in demselben Institut und Review-Team arbeiten, vermuten sie eine deutlich größere Variabilität zwischen verschiedenen Forschungseinrich-

tungen. Als Erklärung für die große Spannweite in den Übereinstimmungen zwischen Bewertern diskutieren die Autoren den Bedarf für verständliche und detaillierte Ausfüllhinweise, um die Reliabilität zu verbessern [26]. Wie der Abbildung 1 zu entnehmen ist, führen die unterschiedlichen Instrumente zu unterschiedlichen Expertenbewertungen derselben Studie. Grund dafür können die verschiedenen Domänen sein, zusätzliche Informationen, die durch manche Instrumente erhoben werden, aber auch die unterschiedliche Tiefe von Fragen. Es mag zu diskordanten Schlussfolgerungen führen, wenn gewisse Aspekte durch manche Instrumente nicht abgefragt werden, anstelle dessen aber andere. Um diese Unsicherheit zu berücksichtigen, wurde eine weniger enge Definition von Übereinstimmung in der Weise vorgenommen, dass eine Bewertung +/- einen Punkt als Übereinstimmung gewertet wurde. Selektionsbias in Form von Publi-

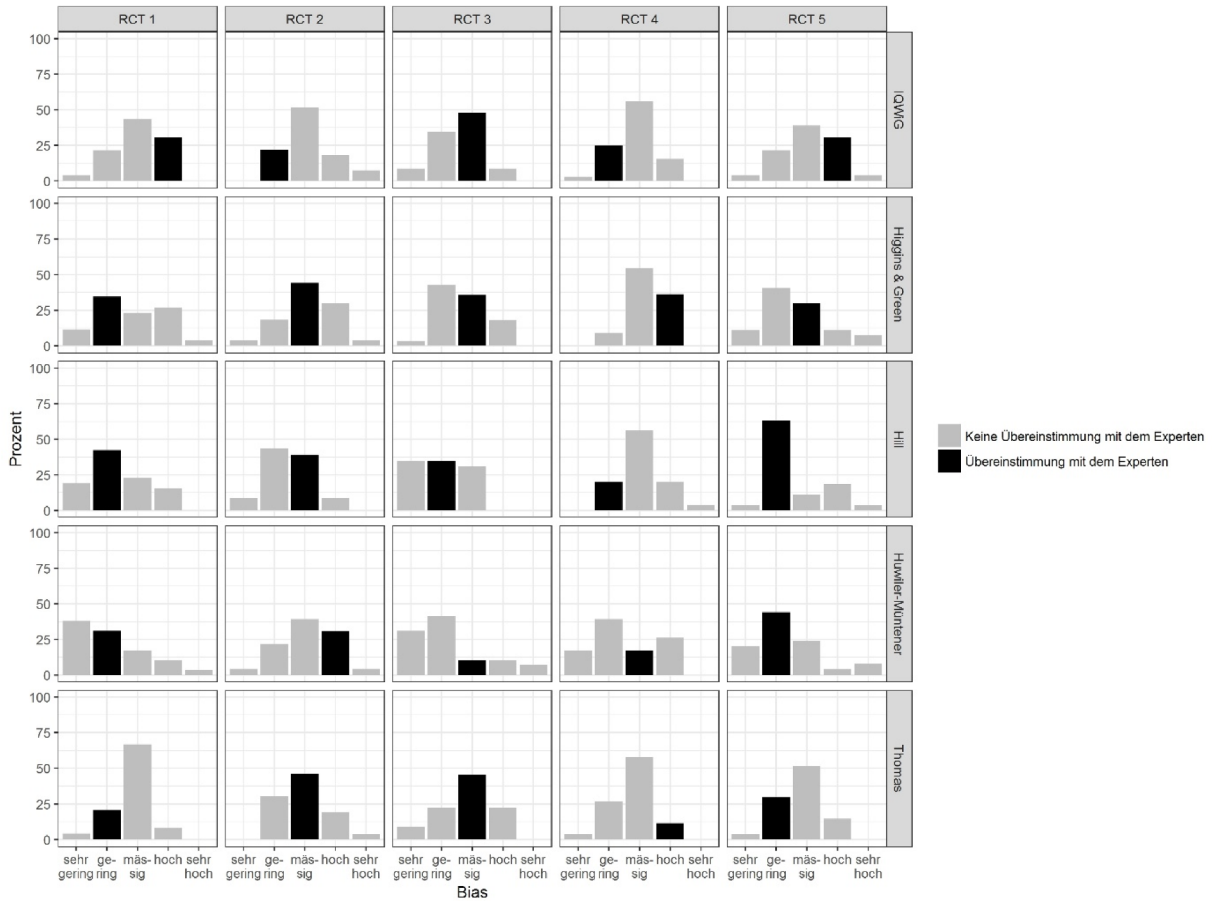


Abbildung 1: Bewertungen von Medizinstudenten nach RCT und Instrument. Balken stellen prozentuale Anteile dar. RCT 1 [21], RCT 2 [22], RCT 3 [27], RCT 4 [23], RCT 5 [34]

Tabelle 4: Übereinstimmung zwischen Studentenbewertungen und Expertenbewertungen

Übereinstimmung +/- 1 Punkt auf Likert-Skala n/N (%)			RCT				
			RCT 1 [21]	RCT 2 [22]	RCT 3 [27]	RCT 4 [23]	RCT 5 [34]
Bewertungsinstrument	Higgins & Green [16]	Medizinstudenten	18/26 (69.23%)	25/27 (92.59%)	27/28 (96.43%)	20/22 (90.91%)	22/27 (81.48%)
		MedMan-Studenten	4/6 (66.67%)	4/4 (100.00%)	3/3 (100.00%)	5/5 (100.00%)	5/5 (100.00%)
	Hill [17]	Medizinstudenten	22/26 (84.62%)	21/23 (91.30%)	26/26 (100.00%)	19/25 (76.00%)	21/27 (77.78%)
		MedMan-Studenten	4/5 (80.00%)	3/3 (100.00%)	5/5 (100.00%)	4/6 (66.67%)	4/4 (100.00%)
	Huwiler-Müntener [18]	Medizinstudenten	25/29 (86.21%)	17/23 (73.91%)	18/29 (62.07%)	19/23 (82.61%)	22/25 (88.00%)
		MedMan-Studenten	4/5 (80.00%)	4/5 (80.00%)	2/6 (33.33%)	4/5 (80.00%)	4/4 (100.00%)
	IQWiG [19]	Medizinstudenten	17/23 (73.91%)	20/27 (74.01%)	21/23 (91.30%)	27/32 (84.38%)	17/23 (73.91%)
		MedMan-Studenten	0/5 (0.00%)	5/5 (100.00%)	3/5 (60.00%)	4/4 (100.00%)	3/6 (50.00%)
	Thomas [20]	Medizinstudenten	22/24 (91.67%)	25/26 (96.15%)	20/22 (90.91%)	18/26 (69.23%)	23/27 (85.19%)
		MedMan-Studenten	5/5 (100.00%)	5/5 (100.00%)	5/5 (100.00%)	2/3 (66.67%)	3/4 (75.00%)

MedMan: Medizinmanagement, IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Bewertung des Biasrisikos auf Likert-Skala: (1) sehr gering, (2) gering, (3) mäßig, (4) hoch, (5) sehr hoch

kationsbias und Berichtsbias kann für alle oben beschriebenen Studien angenommen werden, in denen eindeutig positive Effekte von EBM-Kursen festgestellt wurden.

Das Ausmaß und die Verständlichkeit, mit der einzelne Punkte in den Publikationen beschrieben waren, variierte stark und stellte zum Teil eine unlösbare Aufgabe für unerfahrene nicht graduierte Studenten dar, die auch zu einer Verzerrung der Analyse geführt haben kann.

Es ist bemerkenswert, dass die Beschreibung mancher Aspekte in höher rangigen Zeitschriften (RCT 3 and 4 [27], [23]), die die Einhaltung spezieller Statements zur Berichtsqualität erfordern, für die Studenten weniger verständlich waren als die in Zeitschriften mit einem geringeren Einfluss. Auf die Tatsache, dass eine kritische Bewertung immer auch Subjektivität durch Interpretation beinhaltet, genauso wie auch Scoringsysteme lediglich objektiv erscheinen, aber eine explizite oder implizite Gewichtung ohne empirische Basis [28] bedeuten, sei an dieser Stelle hingewiesen.

Die Anzahl von fünf verschiedenen Bewertungsinstrumenten mag zu hoch gewesen sein. Andererseits führte die Wiederholung von Begriffen auch zu größerer Vertrautheit, und es wurde lediglich Schritt 3 des EBM-Konzepts detailliert behandelt, wodurch die Wissensleistung stark fokussiert war. Insbesondere kann das Trainieren von Schritt 1, der Formulierung einer Forschungsfrage, die operationalisiert werden kann, für nicht graduierte Studenten ohne Erfahrung in wissenschaftlichem Arbeiten viel Zeit beanspruchen. Dasselbe gilt für Schritt 2, der Literaturrecherche in elektronischen Datenbanken.

Das von uns gewählte separate Unterrichten der kritischen Bewertung ist weit verbreitet und auch bekannt als „Journal Club“, für den Teilnehmer unter der Anleitung eines Experten Artikel kritisch lesen und bewerten [4]. Um auf dem neuesten Stand zu bleiben, müssen klinisch tätige Ärzte täglich zahlreiche Artikel lesen, und um das bewältigen zu können, ist ein Training erforderlich [29]. Aus diesem Grund und aufgrund von Empfehlungen dieses Formats für das Unterrichten von EBM unter bestimmten Bedingungen [10] wurde Schritt 3 des EBM-Konzepts in den Fokus genommen. Allerdings wurden dadurch andere Schwierigkeiten hervorgerufen, denn kritische Bewertung schließt Kenntnisse in Epidemiologie, Informatik und Biostatistik ein [9].

Obwohl der Besuch von Vorlesungen in Epidemiologie und Psychologie, in denen anteilig auch Statistik gelehrt wird, in den ersten beiden Studienjahren in Deutschland verpflichtend ist, berichteten 69% unserer Medizinstudenten über schwache oder geringe Kenntnisse in Statistik und/oder Epidemiologie, was zeigt, dass Einstellungen gegenüber einer Sache und Kenntnisse nicht in gleicher Weise verteilt sind, und dass das Unterrichten von EBM die Bedürfnisse unterschiedlich Lernender berücksichtigen muss [8]. Möglicherweise sind Schritt 1 und 2 des EBM-Konzepts für eine Einführung und zur Förderung einer wissenschaftlichen Denkweise besser geeignet. Als Alternative dazu und auch, um dem Vorwurf einer Loslösung von klinischer Praxis zu begegnen [30], kann das Unterrichten von Schritt 4, der Anwendung der Evidenz

auf einzelne Fälle, in Betracht gezogen werden. Allerdings ist das bei nicht Graduierten, denen praktische Erfahrung fehlt und für die das konkrete Ziel nur eine vage Idee darstellt, kaum sinnvoll. Diese Tatsache wird durch langjährige Erfahrungen aus Duke und Stanford unterstrichen, die zu einem Curriculum führten, in dem klinisches Training Bedingung für Forschungserfahrungen ist, weil Studenten dadurch besser darauf vorbereitet sind, das klinische und translationale Potential ihrer Forschungsprojekte zu verstehen [31]. Um Unvoreingenommenheit und eine bessere Bewertung zu erreichen, wäre es ebenfalls günstig, wenn Studenten mit dem ganzen Entwicklungsprozess einer klinischen Studie und der Bedeutung seiner einzelnen Aspekte vertraut wären [32].

Die Fähigkeit, medizinische Literatur zu identifizieren und kritisch zu bewerten, ist unabdingbare Voraussetzung dafür, die beste verfügbare Evidenz bei klinischen Entscheidungen anwenden zu können [13]. Ohne Hintergrundwissen in Methodik und Statistik sind Ärzte einem hohen Risiko ausgesetzt, Evidenz falsch zu interpretieren, wodurch Behandlungsfehler und unerwünschte Ereignisse hervorgerufen werden [28].

Swift et al. [6] untersuchten die Ansichten von 130 Ärzten über Aus- und Weiterbildungsmaßnahmen in Statistik und deren Bedarf für die tägliche Praxis. Als Studenten hatten mehr als die Hälfte der Befragten (60%) den Wert des Fachs für die ärztliche Arbeit unterschätzt, wohingegen die Mehrheit (73%) die Bedeutung im Laufe ihrer Karriere erkannt hatte. Entgegen der wachsenden Überzeugung hinsichtlich der Relevanz von EBM liegt Evidenz für weiterhin bestehende Wissenslücken bei grundlegenden statistischen Konzepten unter praktizierenden Ärzten und in der Medizin forschenden Wissenschaftlern vor [32]. Ebenfalls fehlt es bei einer Mehrheit des Gesundheitspersonals, inklusive von Ärzten, translationalen Forschern, Pflegepersonal und anderen Gesundheitsfachberufen, an Basiswissen über methodologische Schlüsselbegriffe und Quellen der EBM [12], [13]. Um diese Situation zu verbessern, müssen insbesondere Studenten dabei unterstützt werden, die Überzeugung von der Bedeutung dieser Fächer für die klinische Praxis gewinnen zu können [32]. Es bleibt zu hoffen, dass zukünftige Ärzte Forschungsergebnisse besser bewerten können und dazu beitragen, den klinischen Bereich durch Forschung voranzubringen [33].

Danksagung

Wir danken Tobias Goeke, Angelika Gohlke, Anja Hagen, Beate Lux und Monika Nothacker für ihre Experten-Bewertungen.

Ethische Genehmigung

Inhalt und Struktur des Seminars waren durch den Dekan genehmigt. Es gab keinen Kontakt zu Patienten.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Anhänge

Verfügbar unter

<http://www.egms.de/en/journals/zma/2018-35/zma001171shtml>

1. Anhang_1.pdf (58 KB)

Glossar mit Begriffen für Schnellsuche

Literatur

- Godlee F. Milestones on the long road to knowledge. *BMJ*. 2007;344:s2. DOI: 10.1136/bmj.39062.570856.94
- Meskó B. Medical Milestones on the long road to knowledge. [Internet]. 2007. [cited 2015 Sept 09]. Zugänglich unter/available from <http://scinceroll.com/2007/01/18/medical-milestones-on-the-long-road-to-knowledge/>
- Khan KS, Awonuga A, Dwarakanath LS, Taylor R. Assessments in evidence-based medicine workshops: loose connection between perception of knowledge and its objective assessment. *Med Teach*. 2001;23(1):92-94. DOI: 10.1080/01421590150214654
- Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *Can Med Assoc J*. 1998;158(2):177-181.
- Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart T. Teaching critical appraisal skills in healthcare settings. *Cochrane Database Syst Rev*. 2011;(11):CD001270. DOI: 10.1002/14651858.CD001270pub2
- Swift L, Miles S, Price GM, Shepstone L, Leinster SJ. Do doctors need statistics? Doctors' use of and attitudes to probability and statistics. *Stat Med*. 2009;28(15):1969-1981. DOI: 10.1002/sim.3608
- Altman DG, Bland JM. Improving doctors' understanding of statistics. *J Royal Stat Soc*. 1991;154(2):223-267. DOI: 10.2307/2983040
- Straus SE, Green ML, Bell DS, Badgett R, Davis D, Gerrity M, Ortiz E, Shaneyfeldt TM, Whelan C, Mangrulkar R; Society of General Internal Medicine Evidence-Based Medicine Task Force. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ*. 2004;329(7473):1029-1032. DOI: 10.1136/bmj.329.7473.1029
- Ilic D, Maloney S. Methods of teaching medical trainees evidence-based medicine: a systematic review. *Med Educ*. 2014;48(2):124-135. DOI: 10.1111/medu.12288
- Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicines changes anything? A systematic review. *BMJ*. 2004;329(7473):1017. DOI: 10.1136/bmj.329.7473.1017
- Ghali WA, Saitz R, Eskew AH, Gupta M, Quan H, Hershman WY. Successful teaching in evidence-based medicine. *Med Educ*. 2000;34(1):18-22. DOI: 10.1046/j.1365-2923.2000.00402.x
- Young JM, Glasziou P, Ward JE. General practitioners' self rating of skills in evidence based medicine: validation study. *BMJ*. 2002;324(7343):950-951. DOI: 10.1136/bmj.324.7343.950
- Ugolini D, Casanova G, Ceppi M, Mattei F, Neri M. Familiarity of physicians, translational researchers, nurses, and other health professionals with evidence-based medicines terms and resources. *J Canc Educ*. 2014;29(3):514-521. DOI: 10.1007/s13187-014-0631-0
- Weberschock TB, Ginn TC, Reinhold J, Strametz R, Krug D, Bergold M, Schulze J. Change in knowledge and skills of Year 3 undergraduates in evidence-based medicine seminars. *Med Educ*. 2005;39(7):665-671. DOI: 10.1111/j.1365-2929.2005.02191.x
- Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Schriftenreihe Health Technology Assessment, Bd. 102. Köln: Deutsches Institut für Medizinische Dokumentation und Information (DIMDI); 2010.
- Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions 5.1.0*. Oxford, UK: The Cochrane Collaboration; 2011.
- Hill CL, La Valley MP, Felson DT. Secular changes in the quality of published randomized clinical trials in Rheumatology. *Arthritis Rheumatism*. 2002;46(3):779-784. DOI: 10.1002/art.512
- Huwiler-Müntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002;287(21):2801-2804. DOI: 10.1001/jama.287.21.2801
- IQWiG. Früherkennungsuntersuchung von Sehstörungen bei Kindern bis zur Vollendung des 6. Lebensjahres. Abschlussbericht. Nr. 32. Köln: IQWiG-Berichte; 2008.
- Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldview Evidence-based Nurs*. 2004;1(3):176-184. DOI: 10.1111/j.1524-475X.2004.04006.x
- Balegar VK, Kluckow M. Furosemide for packed red cell transfusion in preterm infants: a randomized controlled trial. *J Pediatr*. 2011;59(6):913-918. DOI: 10.1016/j.jpeds.2011.05.022
- Vlug MS, Wind J, Hollmann MW, Ubbink DT, Cense HA, Engel A, Gerhards MF, van Wagenveld BA, van der Zaag ES, van Geloven AA, Sprangers MA, Cuesta MA, Bemelman WA, LAFA study group. Laparoscopy in combination with fast track multimodal management is the best perioperative strategy in patients undergoing colonic surgery. *Ann Surg*. 2011;254(6):868-875. DOI: 10.1097/SLA.0b013e31821fd1ce
- Mega JL, Braunwald E, Wiviott SD, Bassand JP, Bhatt DL, Bode C, Burton P, Cohen M, Cook-Bruns M, Fox KAA, Goto S, Murphy SA, Plotnikov AN, Schneider D, Sun X, Verheugt FW, Gibson CM. ATLAS ACS 2-TIMI 51 Investigators. *N Engl J Med*. 2012;366(1):9-19. DOI: 10.1056/NEJMoa1112277
- Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ*. 2002 325:1338-1341. DOI: 10.1136/bmj.325.7376.1338
- Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. Validity and inter-rater reliability testing of quality assessment instruments. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021-I.) AHRQ Publication No. 12-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012. Zugänglich unter/available from: www.effectivehealthcare.ahrq.gov/reports/final.cfm

26. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012. DOI: 10.1136/bmj.b4012
27. Martins N, Morris, Kelly PM. Food incentives to improve completion of tuberculosis treatment: randomized trial in DILI, Timor-Leste. *BMJ*. 2009;339:b4248. DOI: 10.1136/bmj.b4248
28. Buchberger B, von Elm E, Gartlehner G, Huppertz H, Antes G, Wasem J, Meerpohl JJ. Assessment of risk of bias in controlled studies. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz*. 2014;57(12):1432-1438. DOI: 10.1007/s00103-014-2065-6
29. Kulier R, Gee H, Khan K. Five steps from evidence to effect: exercising clinical freedom to implement research findings. *BJOG*. 2008;115:1197-1202. DOI: 10.1111/j.1471-0528.2008.01821.x
30. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, Germar MJ, Pisake L, Mittal S, Pattinson R, Wolomy-Molondo JJ, Bergh AM, May W, Souza JP, Koppenhoefer S, Khan KS. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training. *JAMA*. 2012;308(21):2218-2225. DOI: 10.1001/jama.2012.33640
31. Laskowitz DT, Drucker RP, Parsonnet J, Cross PC, Gesundheit N. Engaging students in dedicated research and scholarship during medical school: the long term experiences at Duke and Stanford. *Acad Med*. 2010;85(3):419-428. DOI: 10.1097/ACM.0b013e3181ccc77a
32. Miles S, Price GM, Swift L, Shepstone L, Leinster SJ. Statistics teaching in medical school: opinion of practicing doctors. *BMC Med Educ*. 2010;10:75. DOI: 10.1186/1472-6920-10-75
33. Vereijken MW, Kruidering-Hall M, de Jong PG, de Beaufort AJ, Dekker FW. Scientific education early in the curriculum using a constructivist approach on learning. *Perspect Med Educ*. 2013;2(4):209-215. DOI: 10.1007/s40037-013-0072-1
34. Merenstein D, Murphy M, Fokar A, Hernandez RK, Park H, Nsouli H, Sanders ME, Davis BA, Niborski V, Tondou F, Shara NM. Use of fermented dairy probiotic drink containing lactobacillus casei (DN-114 001) to decrease the rate of illness in kids: the DRINK study. A patient-oriented, double-blind, cluster-randomized, placebo-controlled, clinical trial. *Eur J Clin Nutr*. 2010;64(7):669-677. DOI: 10.1038/ejcn.2010.65

Korrespondenzadresse:

B. Buchberger

Universität Duisburg-Essen, Fakultät für
Wirtschaftswissenschaften, Lehrstuhl für
Medizinmanagement, Thea-Leymann-Str. 9, 45127 Essen,
Deutschland, Tel.: +49 (0)201/183-4075, Fax: +49
(0)201/183-4073

barbara.buchberger@medma.uni-due.de

Bitte zitieren als

Buchberger B, Mattivi JT, Schwenke C, Katzer C, Huppertz H, Wasem J. Critical appraisal of RCTs by 3rd year undergraduates after short courses in EBM compared to expert appraisal. GMS J Med Educ. 2018;35(2):Doc24.

DOI: 10.3205/zma001171, URN: urn:nbn:de:0183-zma0011712

Artikel online frei zugänglich unter<http://www.egms.de/en/journals/zma/2018-35/zma001171.shtml>**Eingereicht:** 21.03.2017**Überarbeitet:** 25.11.2017**Angenommen:** 31.01.2018**Veröffentlicht:** 15.05.2018**Copyright**

©2018 Buchberger et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.