

Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF)

Abstract

Objective: The evaluation of medical research performance is a key prerequisite for the systematic advancement of medical faculties, research foci, academic departments, and individual scientists' careers. However, it is often based on vaguely defined aims and questionable methods and can thereby lead to unwanted regulatory effects. The current paper aims at defining the position of German academic medicine toward the aims, methods, and consequences of its evaluation.

Methods: During the Berlin Forum of the Association of the Scientific Medical Societies in Germany (AWMF) held on 18 October 2013, international experts presented data on methods for evaluating medical research performance. Subsequent discussions among representatives of relevant scientific organizations and within three ad-hoc writing groups led to a first draft of this article. Further discussions within the AWMF Committee for Evaluation of Performance in Research and Teaching and the AWMF Executive Board resulted in the final consented version presented here.

Results: The AWMF recommends modifications to the current system of evaluating medical research performance. Evaluations should follow clearly defined and communicated aims and consist of both summative and formative components. Informed peer reviews are valuable but feasible in longer time intervals only. They can be complemented by objective indicators. However, the Journal Impact Factor is not an appropriate measure for evaluating individual publications or their authors. The scientific "impact" rather requires multidimensional evaluation. Indicators of potential relevance in this context may include, e.g., normalized citation rates of scientific publications, other forms of reception by the scientific community and the public, and activities in scientific organizations, research synthesis and science communication. In addition, differentiated recommendations are made for evaluating the acquisition of third-party funds and the promotion of junior scientists.

Conclusions: With the explicit recommendations presented in the current position paper, the AWMF suggests enhancements to the practice of evaluating medical research performance by faculties, ministries and research funding organizations.

Christoph
Herrmann-Lingen¹
Edgar Brunner²
Sibylle Hildenbrand³
Thomas H. Loew⁴
Tobias Raupach⁵
Claudia Spies⁶
Rolf-Detlef Treede⁷
Christian-Friedrich Vahl⁸
Hans-Jürgen Wenz⁹

1 Department of Psychosomatic Medicine and Psychotherapy, University of Göttingen Medical Center, Göttingen, Germany

2 Institute for Medical Statistics, University of Göttingen Medical Center, Göttingen, Germany

3 Institute of Occupational and Social Medicine and Health Services Research, University Hospital of Tübingen, Tübingen, Germany

4 Department of Psychosomatics, University Hospital of Regensburg, Regensburg, Germany

5 Department of Cardiology and Pneumology, University of Göttingen Medical Center, Göttingen, Germany

6 Department for Anesthesiology and Surgical Intensive Care, Charité Berlin, Berlin, Germany

7 CBTM Neurophysiology, Medical Faculty Mannheim of Heidelberg University, Mannheim, Germany

8 Department of Cardiothoracic and Vascular Surgery, University of Mainz Medical Center, Mainz, Germany

1 Status quo

As early as in 1999 the AWMF (Association of the Scientific Medical Societies in Germany) has critically commented on the inappropriate use of methods such as the unadjusted journal impact factor for evaluating medical research performance [1]. These comments have been widely received in German academic medicine and partly been followed by many faculties of medicine [2]. Also from the perspective of individual medical (e.g., [3]) and non-medical disciplines (e.g., [4]), the dominance of journal impact factors in evaluating the performance of individual researchers has been criticized repeatedly. Nevertheless, methodologically questionable quality indicators still play an important role when it comes to evaluating research performance of individuals and institutions, as recently stated in an editorial in *Science* [5]. Based on the San Francisco Declaration on Research Assessment (DORA) there is a growing opposition against the allocation of public resources and the daily practice to decide about scientific careers mainly on the basis of cumulated journal impact factors. Peter Higgs, the current Nobel laureate in physics, recently criticized in an interview with the *Guardian*: “Today I wouldn’t get an academic job... I don’t think I would be regarded as productive enough” (<http://www.theguardian.com/science/2013/dec/06/peter-higgs-interview-underlying-incompetence>).

2 Recommendations

The AWMF makes the following recommendations concerning the evaluation of medical research performance by faculties, ministries of research and research funding organizations:

- The evaluation of medical research performance should be based on a priori and explicitly phrased and communicated aims.
- Informed peer review procedures are particularly useful for evaluating medical research performance. However, because of the high effort involved in such peer review, it appears feasible only at longer time intervals.
- The most important parameter of evaluation is the importance of research for the advancement of scientific medicine or a particular medical discipline.
- For this purpose, the journal impact factor is not an appropriate measure. Therefore it shall not be used for evaluating the research performance of individuals or institutions. It should rather be replaced, as soon as possible, with more appropriate indicators, such as adequately normalized citations rates.
- Besides the reception by the scientific community, also the usefulness for the practice of medicine (e.g., guideline relevance, transfer into practice) or the society as a whole (e.g., disease prevention, economic impact) are considered appropriate indicators of scientific impact in medicine.
- In view of increasing problems with attracting junior scientists and physicians, adequate measures to attract and support young academics make up a second highly important parameter of evaluation.
- The structure and processes of undergraduate academic teaching, measures to support postgraduate junior scientists, and the respective results of these measures are considered appropriate indicators for successful promotion of young scientists. *[A separate position paper of the AWMF and the MFT (German Association of Medical Faculties) will address issues of evaluation in curricular medical teaching and will be issued in the near future.]*
- Depending on the aims of the evaluation, attracted or disbursed third party funds can also be used as parameters for evaluation.
- When evaluating attracted third party funds, public grants or comparable funds based on independently peer-reviewed grant proposals shall receive a higher score than funds from other sources, especially those without a competitive review process.
- Besides simply adding the total amounts of third party funds, the scientific “yield” per sum of money spent should be considered. Suitable algorithms should be developed for defining this ratio.
- Suitable indicators should also be developed for the evaluation of research performed in larger, typically interdisciplinary groups, such as research consortia and multi-author publications. These indicators should take both the individual contribution and the achievement of the group as a whole (added value by networking, coordination etc.) into account. This refers to the scientific impact as well as to jointly attracted third party funds.

3 Rationale for the recommendations

3.1 Overarching aspects of evaluation

Three overarching aspects of evaluation can be identified with regards to medical research performance:

- Aims of the evaluation: optimization of research performance by means of the regulating effects of summative and formative evaluation on different levels (evaluation of individual researchers versus evaluation of institutes, clinical departments, centers or entire faculties)
- Methods of evaluation: dimensions (input/output) [*The term “output” as used here covers both the “impact” of research and promotion for junior scientists.*] and instruments (e.g., informed peer review, metrics)
- Consequences of the evaluation: material and immaterial appraisal and reward for good performance, adaptation of general research environments etc.

From the perspective of the AWMF, evidence base, transparency and acceptance of evaluations are pivotal prerequisites for their success. Evaluations should not be considered acts of top-down control but rather as interactive processes for quality assurance and development of science, and fair allocation of limited research resources and career opportunities. The methods of evaluation follow its defined aims. Conversely, the methods define, what consequences can plausibly be based on the evaluation. Aims and consequences therefore create a necessary framework for methods of evaluation. The main focus of the current paper will be placed on these methods.

3.2 Aims of evaluation

The evaluation of medical research performance should always occur under an a priori clearly defined aim. The aim of the evaluation determines the type, intensity and frequency of evaluations. Evaluation of the quality of research performance is the most critical component in this context. Depending on the precise aims, quality can be operationalized as progress in scientific knowledge or the benefit of research for patient care, undergraduate, postgraduate and continued medical education (including attraction and promotion of junior scientists) or for other societal aims (e.g., prevention, ethical issues, economic relevance). In contrast to the summative evaluations predominating so far (i.e. evaluations of research results) the relative weight of formative evaluations (i.e. evaluation used for optimizing scientific processes) should be increased. Such formative evaluations can serve for giving constructive feedback with the aim of advancing individual careers, scientific programs and institutions and supporting the implementation of good scientific practice.

3.3 Methods of evaluation

Informed peer review procedures such as those that have been endorsed and performed by the German Council of Science and Humanities (“Wissenschaftsrat”) are particularly suitable for evaluating medical research performance. Because of the high effort required for this method (including, e.g., burden on reviewers) informed peer reviews appear feasible solutions only for select purposes (e.g., evaluation of whole faculties, appointment procedures) and at larger time intervals.

Therefore, less costly evaluation methods must also be available. In this context, quantitative parameters can be applied under the precautions mentioned below. However, using an automatic link between certain numeric scores and subsequent (e.g. financial or career-related) consequences is strongly discouraged. Metric indicators should rather serve for informed discussions between the evaluating bodies and the evaluated researchers or institutions. Metric indicators should be interpreted in the context of the specific background of the particular medical discipline evaluated, its research culture, local conditions etc.

In addition to formative evaluations, summative evaluations can be applied for specific regulatory processes in research funding or in preparation for decisions on career development and promotion. These evaluations shall, however, be used with a good sense of proportion and never without a critical appraisal of their regulatory effects, in order to counteract unwanted regulatory effects (e.g., effects on quantity instead of quality [6]; inappropriate overestimation of mainstream research, which tends to produce higher summative scores and may therefore be favored over innovative research). Any unwanted effects should lead to immediate modification of the summative parameters used. Any excessive “evaluatis” is disapproved by the AWMF. This includes, e.g., evaluations that are (in total or in parts) performed without a clear aim or clear consequences, an excessively high frequency of evaluations without appropriate increase in information, or evaluations at time intervals that are too short for initiating meaningful regulatory processes in research planning.

The evaluation of medical research performance shall mainly focus on three core areas:

- The “impact” of research activities in a broader sense, i.e. their contribution to scientific, medical/clinical and other societal progress
- The “input”, i.e. especially the performance in generating competitive third party funds
- The “attraction and promotion of young scientists” as a crucial factor of sustainability

These areas conform with recent recommendations on future research rankings issued by the German Council of Science and Humanities [7]. From the perspective of the AWMF also the criteria additionally proposed by the Council of Science and Humanities, i.e. science transfer,

knowledge transfer and reputation can be subsumed under these three dimensions.

3.3.1 Evaluation of “impact”

As a signer of the San Francisco Declaration on Research Assessment (DORA: <http://am.ascb.org/dora/>) the AWMF commits itself to the requirements of evaluation of publications as stated there. The following aspects are of particular relevance in the context of evaluation of research performance:

- “Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.”
- The scientific content of a publication is much more important than publication metrics or names of journals. This is of particular importance when evaluating junior scientists.
- Funding agencies and institutions should not only consider research publications but also the value of other research outcomes (data sets, software, patents etc.) and keep in mind a broader, also qualitative spectrum of impact measures including influence on politics and medical practice.

When evaluating the impact of an individual’s research performance, it is the position of the AWMF that the core question is whether this individual has contributed to progress in his or her discipline.

This can be assessed or measured on different levels:

- 1st Level: Evaluation of publications
 - a) In recognized scientific journals with peer review
 - b) In other media (books, guidelines etc.)
 - c) Citation by guideline recommendations
- 2nd Level: Active contributions to scientific organizations or boards and editorships
- 3rd Level: Leadership in organizing scientific conferences

It is considered difficult to combine these three levels into a single scale, since no useful conversion factors are available. The levels 1b to 3 should rather be considered relevant indicators in their own right and should be used to supplement the level 1a indicators predominating so far, leading to a multidimensional appraisal with separate subdimensions.

Level 1a

In addition to the requirements stated in the DORA, the following points are suggested:

- As far as bibliometric indicators shall be used for summative evaluations, it must previously be made sure that these indicators are evidence-based, transparent and feasible. For testing the evidence base of indicators, the aim of the evaluation must be kept in

mind (e.g. the desired regulatory effects of the evaluation). The simplicity of performance cannot be the main criterion and must not lead to the application of inappropriate instruments. However, also an evaluation that appears useful in terms of contents must remain feasible under the given conditions. For methodologically adequate evaluation, bibliometric expertise must be “purchased”, if necessary. Such services are commercially available.

- Neither the journal impact factor nor the H-index are suitable measures for evaluation of individuals’ research performance. The journal impact factor is a measure for citations to a journal over a relatively short time frame. It is not sufficiently correlated with citation rates of individuals articles, does not take into account the variance in publication cultures among disciplines and must therefore not be used for evaluating individuals and institutions. The H-index has been seen skeptically due to its numerical instability and dependency on age.
- The use of more differentiated bibliometric analyses is therefore preferred [5], especially the use of field and article type-normalized citation rates. *[I.e. the standardization of citation rates of individual articles is based on their respective disciplines (as a reflected in the subject area of the journal in which an article is published) and on the article type: original papers, review articles and letters to the editor are weighted separately due to very different patterns of citation. It still needs to be determined how field normalization can be performed for interdisciplinary publications or publications from cross-sectional research areas as well as from those journals that are listed in the web of science in a category that differs from the German specification of medical disciplines.]*
- Field normalization reflects and adjusts for the differences in publication and citation cultures across disciplines [8], [9], [10]. Article type normalization takes into account the different average citation rates of e.g., original publications and reviews [11]. *[Of course, also systematic reviews, meta-analyses and guidelines must be considered as genuine scientific publications. See also AWMF Statement dated 9 November 2013. http://www.awmf.org/fileadmin/user_upload/Stellungnahmen/Forschung_und_Lehre/AWMF-Resolution_Wiss-Anerkennung-LL-Arbeit.pdf].* The evaluation window for citations should cover several (e.g. five) years, as suggested by bibliometric research findings [11].
- When evaluating individual research performance, the “dos and don’ts” of individual-related bibliometrics according to Glänzel & Wouters [12] should be followed.
- The individual contributions of each author of a publication should be named in a standardized manner in all journals. Courtesy or honorary authorships are unacceptable.

Level 1b

The following contributions should be counted as discrete publication forms with individually determined weights:

- Monographs and book chapters
- Guidelines and health technology assessment (HTA) reports, even when they have neither been published in a book nor in a scientific journal
- Publications of original data, software developments, patents etc. (see DORA)
- Provision of scientific findings to lay persons for non-scientific practice (e.g., patient guidebooks, press releases etc.)

Level 1c

If a publication is cited by guidelines when immediately justifying specific recommendations (“guideline relevance”) this should be considered an appropriate measure of clinical impact of this publication and should be considered separately.

Level 2

Under this heading, activities for intra- and interdisciplinary networking and quality assurance of research should be evaluated. Relevant aspects of particular importance are

- Editorship in scientific journals as a core instrument for disseminating research results
- Active positions on boards, sections and working groups of scientific societies and organizations
- Active involvement in scientific councils of recognized national or international research-funding and science organizations.
- Outstanding scientific reviewer positions (e.g., colleagues of the German Research Fund [DFG])

Level 3

Organization and leadership in scientific meetings is an important medium of research communication and shall be considered as discrete achievement.

3.3.2 Evaluation of “input”

Input-related parameters of research performance can be defined on various levels. Those factors should be preferred that can be directly influenced by the evaluated person or institution:

- General research framework (basic funding, expertise, strategic concepts, proportion of protected time for research, quality of promotion for junior scientists); this can be influenced on the level of faculties or centers and will not be in the main focus of this statement.
- Attraction and effective use of third party funds; this can be influenced by individual researchers.

Parameters for the evaluation of research performance within faculties:

- Explicit and transparent rules should be established what third party funds will be accepted for evaluation and how they will be weighted.
- When weighting third party funds for evaluation, funds granted after an independent review process must be given higher weights than those provided without independent review.
- Also the source of funding should be weighted: public funding and neutral foundations should be given a higher weight than funds from special interest groups or industrial sponsors.
- Contract research is sufficiently honoured by the funds provided and is appropriate for financing preliminary research for preparing competitive grant applications. It does not justify an additional bonus from public sources. However, in the area of applied science, it can be utilized for evaluating individual researchers.
- A fair and transparent evaluation of individual grant money from collaborative research projects and industry-independent multicenter studies should adequately reflect both the achievement of the principal investigator as well as those of the collaboration partners. This is a necessary prerequisite for a culture of scientific collaboration. Eventually general weighting algorithms must be defined for the participants in various types of research consortia and studies (e.g., one third of evaluation weights for the principal applicant, distribution of the remaining two thirds among all collaboration partners or study centers).
- Funds should be weighted according to the number of positions for scientists. The acquisition of expensive equipment should not be counted as scientific quality criterion. In clinical trials, weights based on the amount of case payments can be considered. Accordingly, flat rates can be used for evaluating third party-funded scientific services.
- The cost effectiveness in terms of scientific output (as described under the chapters “impact” and “attraction and promotion of junior scientists”) per funds granted should be considered as a measure for appropriate spending of resources when evaluating performance of researchers and institutions. For this purpose, suitable algorithms have to be developed.

When evaluating medical research, a bonus for systemic, translational and human subject or patient-related research with concrete reference to practical medicine should be introduced.

3.3.3 Evaluation of attraction and promotion of junior scientists

The attraction and promotion of junior scientists from the beginners’ stage up to the independent researcher is a core issue of evaluation.

The guiding principle should be to get students into science at an early stage and to support them in a sustain-

able way until they acquire sufficient competences for promotion to full professor. Target groups of promotions are undergraduate students, doctoral students and post-doctoral fellows as well as associate professors of medicine and dentistry, physicians, dentists and scientists from related areas. They should be taught a multidimensional model of research that includes scientific work itself, its application in practical medicine, lifelong learning, and teaching. This is in accordance with the role of a “scholar” as defined in the CanMEDs model of teaching [13], in which the role of a scholar is defined as a core role of graduates from medical school curricula in accordance with the international outcome framework. In so far, also the support for scientific competence for students going through the basic medical curriculum should be considered an instrument of promotion of junior scientists. Its evaluation, however, will not be addressed here in detail and only under the aspect of relevance for research, since the evaluation of achievements in teaching will be the subject of a separate position paper which is currently being developed between the AWMF and the German Association of Medical Faculties (“Medizinischer Fakultätentag”).

3.3.3.1 Quantitative indicators

Aspects of junior scientist promotion measurable at the level of institutions (faculty, institute, clinical department) include:

- Broadness and depth of measures for junior scientist promotion e.g.,
 - Structured programs for acquainting students with research: exchange programs, curricular (teaching of scholar competencies, compulsory and elective courses) and hypothesis-based offerings (e.g., journal clubs, term or master theses, “how-to” courses), graduate schools/MD/PhD programs [14], clinical/physician scientist programs [15], mentoring programs in interdisciplinary networks
 - Number of students mentored during medical school (as a modular bridge before the start of scientific specialization [16])
 - Amount of protected time guaranteed [17], i.e. dedicated time for research without competing obligations in teaching and patient care for junior scientists in all steps of their careers until promotion to full professorship. Indicators: Time of exemption from clinical duties in percentages of a full employment (per duration of employment), amount of institution-wide conference times for research (per week or month) with percent participation rates of junior scientists.
- Results of measures for junior scientist promotion e.g.
 - Number of graduates from the programs named above
 - Sustainability of programs; indicators may be e.g., career development, publication, attraction of grant money by junior scientists

- Number of tenure track professorships in research and teaching [18]
- Number of appointments of junior researchers to scientific leadership positions or to clinical leadership positions with a minimum of requirement in the three dimensions of scholar competences (application of scientific results, lifelong learning, teaching)

Individually measurable quantitative criteria of junior scientist promotion include

- Number of adequately supported qualification theses (indicator e.g., number of doctoral students per completed doctorate)
- Number of junior staff supported by structured research and funding programs or with leading role in development of evidence based guidelines
- Career development and research success of junior staff from own research group (criteria see above)

3.3.3.2 Qualitative indicators

Indicators that can be measured on the level of the institution (faculty, institute, clinical department) include:

- Availability and integration (both horizontal and vertical) of appropriate measures to promote junior scientists in the different stages of their careers, e.g. structured doctoral programs and sustainable support programs, internal and external peer review and coaching procedures (learning from the best, from common sense to scientific excellence, soft skills development). These programs shall foster interactions between junior scientists and established experts as suggested by the recommendations of the German Council of Science and Humanities [19] and the German Medical Association [20]. Further indicators may be the existence of research tracks and representatives of scientific societies within the institution.
- Availability of quality standards (e.g., good scientific practice) and scientific infrastructure including, among others, programs for startup financing (for e.g., one year) in order to pay young scientists on the way to their first DFG grant, infrastructure directly related to grant applications (among others, courses about good clinical and scientific practice or good laboratory practice, grant counselling, support with related paperwork when dealing with e.g., animal protection authorities, availability of electronic laboratory diaries, use of core facilities or core research units) and clinical trials infrastructure (to support formal requirements from health authorities etc., among others: data protection, ethics committees, Federal Institute for Pharmaceuticals and Medical Products (BfArM), registration with state authorities and trial registries such as clinicaltrials.gov, writing of safety reports, procedures for prepublication of methods, evidence based monitoring of documents).
- Availability of measures to increase transparency and equal opportunities. E.g., transparency of the scientific

profile as an instrument for targeted choice of the university by junior scientists.

- Flat hierarchies with e.g., trainee representation/speakers for assistant researchers, tandem professorships (e.g. following the Swiss example: assistant professorship with 50% time for research and teaching plus 50% for patient care, followed by a tenure track with 70% for research and teaching plus 30% for patient care or reverse); equal payment for scientific and clinical work over the whole course of the career; sustainability by better compatibility of career and family (childcare, eventually also at night, during weekends and with priority for scientists, childcare during meetings, seminars and conferences, emergency childcare).

Individually measurable qualitative criteria include:

- Active participation in the promotion of young scientists by collaborating in e.g., DFG young scientist academies, summer schools for outstanding doctoral students, personal dedication as a model function, teaching of basic scientific competences to group members, e.g., teaching the difference between practice (everyday knowledge) profession (professional knowledge) and science (scientific knowledge) as well as bridging the gaps between analytics, transformation and theory; early integration of junior scientists in working groups with increasing individual responsibilities.
- Quality of research performance of group members, e.g. reproducibility of results of members from the own group by other research groups, self-assessment: evaluation of both the most important results of own research and of the independence of own research by junior scientists as a measure of junior scientists' promotion by their respective mentors, consequent tracking of junior scientists' research, which should be reflected in a line of research and make clear the relevance of this research.

3.4 Consequences of evaluation

Possible consequences of the evaluation have to be clearly defined a priori. They should follow the aims of the evaluation and keep in mind its methodological limitations. Beside immediate feedback and joint discussion of evaluation results they can consist of a targeted application of instruments for organizational (e.g., creation of new research foci), project (e.g., funding decision) and career planning (e.g. appointment, tenure, mentoring), as long as they are based on a well-balanced and transparent procedure.

The performance-related allocation of funds is only one of numerous possible consequences of evaluation and its regulatory effects are seen controversially [21]. Appreciation of good work is of particular importance for the vast majority of scientists who are highly and intrinsically motivated. This should equally cover achievements in research, teaching and (in clinical medicine) also in patient care. The recognition can also be expressed by provision of time resources. In contrast, over the long

term a predominance of financial incentives as extrinsic motivators runs the risk to undermine intrinsic motivation. This is particularly true, when the underlying evaluation processes are perceived as intransparent or unjust. Therefore a sufficient basic funding of institutions is of high importance and should be subject to comprehensive evaluations only in longer time intervals in order to be adapted to new developments.

Notes

Acknowledgement

The present position paper is based on the results of the Berlin Forum of the AWMF held on 18 October 2013 under the headline "Methods for evaluation of medical research performance". At this forum, invited experts, representatives from major science organizations (German Research Fund, Association of Medical Faculties, German Aerospace Center Project Management Agency, German Council of Science and Humanities) and participants of ad-hoc writing committees presented proposals which were subsequently subject to discussions within the AWMF Committee for Evaluation of Performance in Research and Teaching, condensed into a draft version and consented by the executive board of the AWMF. The authors thank all participants for their contributions to the present paper. In particular the following person should be mentioned here (in alphabetical order): Prof. K.-M. Debatin, Prof. R. Deinzer, Prof. W. Glänzel, Prof. C. Graf, Prof. H.-J. Heinze, Prof. S. Hornbostel, Dr. M. Kordel-Bödighheimer, Dr. T. Kostuj, Prof. H. K. Kroemer, Dr. A. Lücke, Prof. P. Meier-Abt, Dr. S. Moritz, Prof. A. F. J. van Raan, Prof. K. Rahn, Prof. G. Theilmeier, Dr. W. Warmuth.

Conflicts of interest

The authors declare the following conflicts of interest: All authors are affiliated with German universities, which includes leadership duties and participation in academic self-government. All authors are also members of scientific societies, some (THL, RDT, CFV) of them in leading positions. CHL, RDT and CS are AWMF board members. All authors are authors of scientific publications. In preparing this manuscript, CHL, SH, RDT and HJW have received travel support from the AWMF and/or scientific societies for attending meetings. CHL, RDT and CS have received research grants from public sources. CHL, SH and RDT have received research grants from private entities. CHL and RDT have received honoraria from private entities. CS has been vice dean for teaching until May 2014, RDT is vice dean for research. CHL, EB, THL, RDT, CFV and CS are editors, associate editors or editorial board members for scientific journals. CHL and EB are book editors. CHL, EB, RDT, CFV and CS are reviewers for research funding organisations or endowments. THL is endowment council member.

Authorship

The co-authors are listed in alphabetical order.

References

- Frömter E, Brähler E, Langenbeck U, Meenen NM, Usadel KH. Das AWMF-Modell zur Evaluierung publizierter Forschungsbeiträge in der Medizin [The AWMF model for the evaluation of published research papers in medicine. Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften (Working Group of the Scientific Medical Specialty Societies)]. *Dtsch Med Wochenschr.* 1999 Jul;124(30):910-5.
- Brähler E, Strauss B. Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten : Eine aktuelle Übersicht [Performance-oriented allocations of financial resources at medical schools: an overview]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2009 Sep;52(9):910-6. DOI: 10.1007/s00103-009-0918-1
- Vahl CF. Forschungsbewertung: Fairness für forschende Chirurgen: ein Plädoyer. *Dtsch Arztebl.* 2008;105(12):A-625-8.
- Adler R, Ewing J, Taylor P. Citation Statistics – A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Stat Sci.* 2009;24(1):1-14. DOI: 10.1214/09-STS285
- Alberts B. Impact factor distortions. *Science.* 2013 May;340(6134):787. DOI: 10.1126/science.1240319
- Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* 2014 Jan;383(9913):267-76. DOI: 10.1016/S0140-6736(13)62228-X
- Wissenschaftsrat. Empfehlungen zur Zukunft des Forschungsratings. Drs. 3409-13, Mainz 25.10.2013. Available from: <http://www.wissenschaftsrat.de/download/archiv/3409-13.pdf>
- Van Raan AFJ. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung – Theorie und Praxis.* 2003;12(1):20-9.
- Van Raan AFJ. Statistical Properties of Bibliometric Indicators: Research Group Indicator Distributions and Correlations. *J Am Soc Inf Sci Technol.* 2006;57(3):408-430. DOI: 10.1002/asi.20284
- Schulze N, Michels C, Frietsch R, Schmoch U, Conchi S. 2. Indikatorbericht Bibliometrische Indikatoren für den PFI Monitoring Bericht 2013. Hintergrundbericht für das Bundesministerium für Bildung und Forschung (BMBF). Berlin, Karlsruhe, Bielefeld: Institut für Forschungsinformation und Qualitätssicherung (iFQ), Fraunhofer-Institut für System- und Innovationsforschung ISI und Universität Bielefeld, Institut für Wissenschafts- und Technikforschung (IWT); 2012 [cited 23.6.2014]. Available from: http://www.bmbf.de/pubRD/Indikatorbericht_PFI_2013.pdf
- Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics.* 2006;67(3):491-502. DOI: 10.1556/Scient.67.2006.3.10
- Glänzel W, Wouters P. The dos and don'ts in individual level bibliometrics. Paper presented at the 14th ISSI Conference, Vienna, 15-18 July 2013. Available from: <http://de.slideshare.net/paulwouters1/issi2013-wg-pw>
- Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach.* 2007 Sep;29(7):642-7. DOI: 10.1080/01421590701746983
- Basu Ray I, Henry TL, Davis W, Alam J, Amedee RG, Pinsky WW. Consolidated academic and research exposition: a pilot study of an innovative education method to increase residents' research involvement. *Ochsner J.* 2012;12(4):367-72.
- Gaetgens C. Clinical Scientist – Neue Karrierewege in der Hochschulmedizin. Werkstattgespräch am 27.-28.9.2013 in Schloss Herrenhausen, Hannover. Ergebnisse und Schlussfolgerungen. [cited 20.5.2014]. Available from: http://www.volkswagenstiftung.de/fileadmin/downloads/publikationen/veranstaltungsberichte/Veranstaltungsbericht_Clinical_Scientist.pdf
- Bierer SB, Chen HC. How to measure success: the impact of scholarly concentrations on students—a literature review. *Acad Med.* 2010 Mar;85(3):438-52. DOI: 10.1097/ACM.0b013e3181cccdbd4
- Sullivan R, Badwe RA, Rath GK, Pramesh CS, Shanta V, Digumarti R, D'Cruz A, Sharma SC, Viswanath L, Shet A, Vijayakumar M, Lewison G, Chandu M, Kulkarni P, Bardia MR, Kumar S, Sarin R, Sebastian P, Dhillon PK, Rajaraman P, Trimble EL, Aggarwal A, Vijaykumar DK, Purushotham AD. Cancer research in India: national priorities, global results. *Lancet Oncol.* 2014 May;15(6):e213-22. DOI: 10.1016/S1470-2045(14)70109-3
- Kelley WN, Stross JK. Faculty tracks and academic success. *Ann Intern Med.* 1992 Apr;116(8):654-9. DOI: 10.7326/0003-4819-116-8-654
- Wissenschaftsrat. Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Drs. 1656-11, Halle 11.11.2011. Available from: <http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf>
- Bundesärztekammer. Curriculum Ärztliches Peer Review. 2. Auflage. Berlin: Bundesärztekammer; 2013. (Texte und Materialien der Bundesärztekammer zur Fortbildung und Weiterbildung; 30). Available from: <http://www.bundesaerztekammer.de/downloads/CurrAerztlPeerReview2013.pdf>
- Krempkow R, Schulz P. Welche Effekte hat die leistungsorientierte Mittelvergabe? Das Beispiel der medizinischen Fakultäten Deutschlands. *Hochschule.* 2012;(2):122-42.

Corresponding author:

Prof. Dr. med. Christoph Herrmann-Lingen
Department of Psychosomatic Medicine and
Psychotherapy, University of Göttingen Medical Center,
Von-Siebold-Str. 5, 37075 Göttingen, Germany, Phone:
+49-551-39-6707
cherrma@gwdg.de

Please cite as

Herrmann-Lingen C, Brunner E, Hildenbrand S, Loew TH, Raupach T, Spies C, Treede RD, Vahl CF, Wenz HJ. Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS Ger Med Sci.* 2014;12:Doc11. DOI: 10.3205/000196, URN: urn:nbn:de:0183-0001969

This article is freely available from

<http://www.egms.de/en/journals/gms/2014-12/000196.shtml>

Received: 2014-06-23
Published: 2014-06-26

Copyright

©2014 Herrmann-Lingen et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share – to copy, distribute and transmit the work, provided the original author and source are credited.

AWMF-Positionspapier zur Evaluation der medizinischen Forschungsleistung

Zusammenfassung

Ziele: Die Evaluation medizinischer Forschungsleistung ist eine zentrale Voraussetzung für die gezielte Weiterentwicklung von Fakultäten, Forschungsschwerpunkten, Instituten bzw. Kliniken sowie Wissenschaftlerkarrieren. Sie basiert aber vielfach auf unklaren Zielvorstellungen und fragwürdigen Methoden und kann zu Fehlsteuerungen führen. Ziel der vorliegenden Arbeit ist eine Positionsbestimmung der wissenschaftlichen Medizin gegenüber den Zielen, Methoden und Konsequenzen ihrer Evaluation.

Methoden: Im Rahmen des Berliner Forums der AWMF vom 18.10.2013 wurden von internationalen Experten Befunde zur Methodik der Evaluation medizinischer Forschungsleistung referiert, von Vertretern relevanter Wissenschaftsorganisationen diskutiert und in drei Adhoc-Schreibgruppen in einen Textentwurf gefasst. Dieser wurde mehrfach in der AWMF-Kommission für Leistungsevaluation in Forschung und Lehre sowie im AWMF-Präsidium diskutiert und in der vorliegenden finalen Fassung konsentiert.

Ergebnisse: Die AWMF empfiehlt eine Veränderung des Systems der Bewertung medizinischer Forschungsleistungen. Evaluationen sollten eine a priori klar definierte und kommunizierte Zielsetzung besitzen und neben summativen auch formative Elemente enthalten. Neben sinnvollen aber nur in längeren Intervallen praktikablen informierten peer reviews können auch objektive Indikatoren zur Evaluation eingesetzt werden. Die Verwendung des Journal Impact Factors zur Bewertung einzelner Publikationen oder ihrer Autoren wird dabei abgelehnt. Der wissenschaftliche „Impact“ erfordert vielmehr eine multidimensionale Beurteilung, in der z.B. normierte Zitationsraten wissenschaftlicher Publikationen ebenso Berücksichtigung finden wie andere Formen der Rezeption in Fachwelt und Öffentlichkeit sowie Tätigkeiten im Bereich der Wissenschaftsorganisation, -synthese und -kommunikation. Daneben werden Vorschläge zu einer differenzierten Bewertung von Drittmittelwerbung und Nachwuchsförderung gemacht.

Folgerungen: Mit den konkreten Empfehlungen des vorliegenden Positionspapiers legt die AWMF den Fakultäten, Ministerien und Organisationen der Forschungsförderung eine Weiterentwicklung der Evaluationspraxis medizinischer Forschungsleistungen nahe.

**Christoph
Herrmann-Lingen¹
Edgar Brunner²
Sibylle Hildenbrand³
Thomas H. Loew⁴
Tobias Raupach⁵
Claudia Spies⁶
Rolf-Detlef Treede⁷
Christian-Friedrich Vahl⁸
Hans-Jürgen Wenz⁹**

1 Klinik für Psychosomatische Medizin und Psychotherapie, Universitätsmedizin Göttingen, Göttingen, Deutschland

2 Institut für Medizinische Statistik, Universitätsmedizin Göttingen, Göttingen, Deutschland

3 Institut für Arbeitsmedizin, Sozialmedizin und Versorgungsforschung, Universitätsklinikum Tübingen, Tübingen, Deutschland

4 Abteilung für Psychosomatik, Universitätsklinikum Regensburg, Regensburg, Deutschland

5 Klinik für Kardiologie und Pneumologie, Universitätsmedizin Göttingen, Göttingen, Deutschland

6 Klinik für Anästhesiologie mit Schwerpunkt Operative Intensivmedizin der Charité – Universitätsmedizin Berlin, Berlin, Deutschland

7 CBTM, Lehrstuhl Neurophysiologie, Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Deutschland

8 Klinik für Herz-Thorax und Gefäßchirurgie,

Universitätsmedizin Mainz,
Mainz, Deutschland

9 Klinik für Zahnärztliche
Prothetik, Propädeutik und
Werkstoffkunde, Christian-
Albrechts-Universität Kiel,
Kiel, Deutschland

1 Ausgangslage

Die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) hat bereits im Jahr 1999 kritisch zur Bewertung der medizinischen Forschungsleistung, insbesondere auch zur Verwendung des unadjustierten Journal Impact Factors, Stellung genommen [1]. Diese Empfehlungen wurden breit rezipiert und an vielen Fakultäten zumindest teilweise umgesetzt [2]. Auch aus der Perspektive einzelner medizinischer (z.B. [3]) und nichtmedizinischer Fächer (z.B. [4]) wurde die Dominanz der Journal Impact-Faktoren für die Leistungsbewertung individueller Wissenschaftler wiederholt kritisiert. Dennoch spielen methodisch zumindest fragwürdige Qualitätsindikatoren nach wie vor eine bedeutende Rolle bei der Bewertung der Forschungsleistung von Individuen und Institutionen, wie kürzlich in einem Editorial in *Science* erneut betont wurde [5]. Unter Verweis auf die San Francisco Declaration on Research Assessment (DORA) wird der Verteilung öffentlicher Mittel und der Praxis der Entscheidung über wissenschaftliche Karrieren auf der Basis summierter Journal Impact-Faktoren jetzt auch auf breiter Basis entgegengetreten. Peter Higgs, der aktuelle Nobelpreisträger für Physik, kritisierte kürzlich in einem Interview mit dem *Guardian*: „Today, I wouldn't get an academic job. ... I don't think I would be regarded as productive enough“ (<http://www.theguardian.com/science/2013/dec/06/peter-higgs-interview-underlying-incompetence>).

2 Empfehlungen

Die AWMF empfiehlt den Fakultäten, Wissenschaftsministerien und Institutionen der Forschungsförderung folgende Punkte zur Bewertung der medizinischen Forschungsleistung:

- Der Evaluation medizinischer Forschungsleistung sollte eine jeweils a priori explizit formulierte und kommunizierte Zielsetzung zugrunde liegen.
- Methodisch eignen sich zur Evaluation medizinischer Forschungsleistung insbesondere informierte peer-review-Verfahren. Wegen des damit verbundenen Aufwands stellen diese jedoch nur in größeren zeitlichen Intervallen gangbare Wege dar.
- Wichtigster Parameter der Evaluation ist die Bedeutung der Forschungsleistung für die Weiterentwicklung der wissenschaftlichen Medizin bzw. eines spezifischen Fachgebiets.
- Der Journal Impact Factor ist hierfür kein geeignetes Instrument. Er soll daher nicht für die Bewertung der Forschungsleistung von Individuen oder Institutionen verwendet werden sondern schnellstmöglich durch geeignete Indikatoren, z.B. adäquat normierte Zitationsraten ersetzt werden.
- Geeignete Indikatoren des medizinisch-wissenschaftlichen Impacts umfassen neben der Rezeption in der wissenschaftlichen Fachwelt auch die Nützlichkeit für die praktische Medizin (z.B. Leitlinien-Relevanz, Praxistransfer) bzw. die Gesellschaft insgesamt (z.B. Krankheitsprävention, ökonomischer Nutzen).
- Angesichts zunehmender Nachwuchsprobleme in der medizinischen Forschung und der Medizin insgesamt stellt eine geeignete Nachwuchsgewinnung und -förderung einen zweiten wesentlichen Evaluationsparameter dar.
- Geeignete Indikatoren für erfolgreiche Nachwuchsförderung sollten sowohl die Struktur- und Prozessqualität der akademischen Lehre als auch Maßnahmen zur Förderung des postgraduierten Nachwuchses und ihre jeweiligen Ergebnisse erfassen. [*Eine separate Stellungnahme der AWMF und des MFT (Medizinischer Fakultätentag) zur Evaluation der curricularen Lehre befindet sich gegenwärtig in Vorbereitung.*]
- Je nach Zielsetzung der Evaluation können auch die eingeworbenen bzw. verausgabten Drittmittel als Parameter genutzt werden.
- Bei der Evaluation der Drittmittelinwerbung sollten antragsbasierte und unabhängig begutachtete Mittel öffentlicher Fördermittelgeber oder vergleichbarer Förderorganisationen höher bewertet werden als unbegutachtete Zuwendungen anderer Herkunft.
- Neben einer reinen Bewertung von Drittmittelsummen sollte der wissenschaftliche „Ertrag“ pro eingesetzter Fördersumme Berücksichtigung finden. Hierfür sollten geeignete Algorithmen entwickelt werden.
- Es sind geeignete Indikatoren für die Evaluation in größeren, oft interdisziplinären Gruppen erbrachter Leistungen (z.B. Forschungsverbünde, Vielautoren-Publikationen) zu entwickeln, die sowohl dem individuellen Beitrag als auch der Gruppenleistung insgesamt (Mehrwert durch Vernetzung, Koordination etc.)

Rechnung tragen. Dies betrifft sowohl den wissenschaftlichen Impact als auch gemeinsam eingeworbene Drittmittel.

3 Begründung der Empfehlungen

3.1 Übergeordnete Aspekte der Evaluation

Bei der Evaluation medizinischer Forschungsleistung können drei übergeordnete Aspekte unterschieden werden:

- Ziele der Evaluation: Leistungsoptimierung der Forschung durch die Steuerungswirkung von summativer und formativer Evaluation auf verschiedenen Ebenen (Beurteilung von Einzelpersonen vs. Beurteilung von Instituten, Kliniken, Schwerpunkten oder Fakultäten)
- Methoden der Evaluation: Dimensionen (input/output) [Der Begriff „output“ umfasst hier sowohl den „Impact“ der Forschung als auch die Nachwuchsförderung.] und Instrumente (z.B. informierter peer review, Maßzahlen)
- Konsequenzen der Evaluation: Materielle und immaterielle Bewertung und Belohnung guter Leistung, Anpassung der Rahmenbedingungen etc.

Aus Sicht der AWMF sind Evidenzbasierung, Transparenz und Akzeptanz von Evaluationen wesentliche Voraussetzungen für ihren Erfolg. Evaluationen sollten keine hoheitlichen Akte sondern interaktive Prozesse zur Qualitätssicherung und -entwicklung der Wissenschaft und fairen Allokation begrenzter Forschungsressourcen und Karriereoptionen sein. Dabei folgt die Methodik den Zielsetzungen der Evaluation. Umgekehrt ergibt sich aus ihr, welche Konsequenzen sinnvoller Weise aus der Evaluation zu ziehen sind. Zielsetzungen und Konsequenzen stellen somit notwendige Rahmenbedingungen der Evaluationsmethodik dar, die den Schwerpunkt der vorliegenden Stellungnahme bildet.

3.2 Ziele der Evaluation

Die Evaluation medizinischer Forschungsleistungen sollte jeweils unter einer bereits a priori klar definierten Zielsetzung erfolgen. Das Evaluationsziel bestimmt die Art, Intensität und Häufigkeit von Evaluationen. Entscheidend ist dabei die Beurteilung der Qualität der Forschungsleistung. Diese kann sich je nach konkreter Zielsetzung am erzielten Erkenntnisfortschritt oder dem Nutzen der Forschung für die Krankenversorgung, die Ärzte-Aus-, Fort- und Weiterbildung incl. Nachwuchsgewinnung und -förderung oder für andere gesellschaftliche Ziele (z.B. Prävention, ethische Fragestellungen, ökonomische Relevanz) orientieren.

Gegenüber den bislang vorherrschenden summativen Evaluationen (also Evaluationen zur Ergebnisbewertung) ist das Gewicht formativer Evaluationen (also solcher zur Prozessoptimierung) zu erhöhen. Diese können konstruk-

tives Feedback mit dem Ziel der Weiterentwicklung individueller Karrieren, wissenschaftlicher Programme und Einrichtungen geben und die Implementierung guter wissenschaftlicher Praxis fördern.

3.3 Methoden der Evaluation

Methodisch eignen sich zur Evaluation medizinischer Forschungsleistung insbesondere informierte peer-review-Verfahren wie sie z.B. vom Wissenschaftsrat vertreten und durchgeführt werden. Wegen des damit verbundenen Aufwands (incl. Belastung der Gutachter) stellen diese jedoch nur für ausgewählte Zielsetzungen (z.B. Evaluation von Fakultäten, Berufungsverfahren) und in größeren zeitlichen Intervallen gangbare Wege dar.

Es müssen daher auch weniger aufwändige Evaluationsmethoden zur Verfügung stehen. Hierfür können unter den unten aufgeführten Prämissen auch quantitative Parameter zum Einsatz kommen. Ein Automatismus zwischen dem Erreichen einer bestimmten Maßzahl und einer nachfolgenden (z.B. finanziellen oder karrierebezogenen) Konsequenz wird aber abgelehnt. Vielmehr sollten auch metrische Indikatoren zunächst Gegenstand einer gemeinsamen Erörterung von Evaluierendem und Evaluiertem sein und gemeinsam vor dem spezifischen Hintergrund des Fachgebiets und seiner Forschungskultur, lokaler Besonderheiten etc. bewertet werden.

Neben formativen Evaluationen können summative Evaluationen zur gezielten Steuerung der Forschungsförderung oder Vorbereitung von Karriereschritten zum Einsatz kommen. Ihr Einsatz soll aber mit Augenmaß und niemals ohne kritische Überprüfung ihrer Steuerungswirkung erfolgen, so dass Fehlsteuerungen (z.B. Quantität vor Qualität [6]; Innovationsfeindlichkeit durch Fokussierung der Forschung auf den hinsichtlich summativer Parameter leichter vorhersagbaren Mainstream) frühzeitig durch Änderung der Parameter der summativen Evaluation entgegengewirkt werden kann.

Eine überzogene „Evaluitis“ wird von der AWMF kritisch gesehen. Hiervon ist z.B. auszugehen, wenn Evaluationen ganz oder in Teilen ohne klare Zielsetzung bzw. Konsequenzen erfolgen oder wenn eine zu hohe Dichte der Evaluationen im Missverhältnis zu ihrem Erkenntnisgewinn bzw. zu den für sinnvolle Steuerungsprozesse in der Wissenschaft benötigten Zeitintervallen steht.

Die Evaluation medizinischer Forschungsleistung soll im Wesentlichen drei zentrale Bereiche erfassen:

- den „Impact“ der Forschungstätigkeit im weiteren Sinne, also ihren Beitrag zum wissenschaftlichen, klinisch-medizinischen und sonstigen gesellschaftlichen Fortschritt
- den „Input“, hier insbesondere die Leistung bei der kompetitiven Drittmittelinwerbung
- die Nachwuchsförderung als wesentlichen Nachhaltigkeitsfaktor.

Diese Bereiche stehen in Übereinstimmung mit den aktuellen Empfehlungen des Wissenschaftsrats [7] zu zukünftigen Forschungsratings. Aus Sicht der AWMF lassen sich

unter diese drei Dimensionen auch die vom Wissenschaftsrat zusätzlich vorgeschlagenen Kriterien Wissenstransfer, Wissensvermittlung und Reputation subsumieren.

3.3.1 Evaluation des „Impacts“

Mit dem Beitritt zur San Francisco Declaration on Research Assessment (DORA: <http://am.ascb.org/dora/>) bekennt sich die AWMF zu den dort erhobenen Forderungen zur Publikationsbewertung. Von besonderer Bedeutung im Kontext der Leistungsevaluation sind hier:

- Keine Nutzung von Journal-Metriken wie dem Journal Impact Factor als Qualitätsmaß zur Beurteilung der Qualität individueller Artikel oder ihrer Autoren bzw. deren Einstellung oder (Be-) Förderung.
- Der wissenschaftliche Gehalt einer Arbeit ist wesentlich bedeutsamer als Publikationsmetriken oder Journal-Namen. Dies gilt besonders auch bei der Beurteilung von NachwuchswissenschaftlerInnen.
- Förderer und Institutionen sollen neben Forschungspublikationen auch den Wert anderer Forschungsergebnisse (Datensätze, Software, Patente...) bedenken und ein breiteres, auch qualitatives Spektrum an Impactmaßen incl. Einfluss auf Politik und Praxis beachten.

Kernpunkt bei der Beurteilung des Impacts der Forschungsleistungen einer Person ist aus Sicht der AWMF die Frage, ob diese Person in ihrem Fach etwas bewegt hat.

Dies kann auf verschiedenen Ebenen erfasst bzw. gemessen werden:

- 1. Ebene: Bewertung von Publikationen
 - a) in anerkannten wissenschaftlichen Fachzeitschriften mit peer review
 - b) in anderen Medien (Bücher, Leitlinien etc.)
 - c) Zitation durch Leitlinienempfehlungen
- 2. Ebene: Aktive Aufgaben in Wissenschaftsorganisationen, Gremien und Herausgeberschaften
- 3. Ebene: Organisation/Leitung wissenschaftlicher Tagungen

Dabei ist es methodisch schwierig, die drei Ebenen auf einer gemeinsamen Skala zusammenzufassen, da keine sinnvollen Umrechnungsfaktoren existieren. Die Ebenen 1b bis 3 sind aber zusätzlich zur bislang vorherrschenden Ebene 1a als Indikatoren mit eigener Berechtigung zu berücksichtigen und sollten daher im Sinne einer mehrdimensionalen Bewertung in separaten Rubriken in die Evaluation einfließen.

Ebene 1a

In Ergänzung der Forderungen der DORA werden folgende Punkte angeregt:

- Soweit einer summativen Evaluation bibliometrische Indikatoren zugrundegelegt werden sollen, muss zuvor

sichergestellt werden, dass diese evidenzbasiert, transparent und praktikabel sind. Bei der Prüfung der Evidenzbasierung ist das Evaluationsziel zugrunde zu legen (z.B. gewünschte Steuerungswirkung der Evaluation). Die Einfachheit der Durchführung kann dabei nicht wesentliches Kriterium sein und zur Wahl ungeeigneter Instrumente führen. Dennoch muss eine inhaltlich sinnvolle Evaluation unter den gegebenen Bedingungen durchführbar bleiben. Für eine methodisch adäquate Evaluation muss bei Bedarf die erforderliche bibliometrische Expertise „eingekauft“ werden. Entsprechende Angebote sind auf dem Markt verfügbar.

- Weder der Journal Impact Factor noch der H-Index eignen sich für die individuelle Leistungsbeurteilung. Der Journal Impact Factor stellt ein Maß für die Zitationshäufigkeit einer Zeitschrift über einen relativ kurzen Zeitraum dar. Er korreliert nicht hinreichend mit den Zitationsraten individueller Beiträge, bildet unterschiedliche Publikationskulturen verschiedener Fächer nicht ab und ist daher für die Beurteilung von Individuen und Institutionen ungeeignet. Der H-Index ist wegen numerischer Instabilität und Altersabhängigkeit kritisch zu sehen.
- Die Verwendung differenzierterer bibliometrischer Analysen wird daher bevorzugt [5], insbes. die Verwendung Feld- und Artikeltyp-normierter Zitationsraten. *[D.h., die Standardisierung der Zitationsraten einzelner Artikel erfolgt auf der Basis des jeweiligen Fachgebiets (ermittelt über die Fachgebietszuordnung der Zeitschrift, in der der Artikel erscheint) sowie des Artikel-Typs: Originalarbeiten, Übersichten und Leserbriefe werden wegen sehr unterschiedlicher Zitationsmuster separat gewertet. Offen ist noch, wie die Fachzugehörigkeit interdisziplinärer Publikationen bzw. von Publikationen aus Querschnittsbereichen oder aus solchen Journalen bewertet werden soll, deren Zuordnung im Web of Science von dem deutschen Fächerkanon abweicht.]* Die Feldnormierung reflektiert und adjustiert für die unterschiedlichen Publikations- und Zitationskulturen verschiedener Fachgebiete [8], [9], [10]. Die Artikeltyp-Normierung trägt den unterschiedlichen mittleren Zitationsraten von z.B. Originalarbeiten und Übersichten Rechnung [11]. *[Selbstverständlich sind auch systematische Übersichtsarbeiten, Metaanalysen und Leitlinien als originäre wissenschaftliche Publikationsleistungen zu werten, vgl. Resolution der AWMF vom 9.11.2013 http://www.awmf.org/fileadmin/user_upload/Stellungnahmen/Forschung_und_Lehre/AWMF-Resolution_Wiss-Anerkennung-LL-Arbeit.pdf].* Als Bewertungsfenster sollen angesichts der empirischen Datenlage zum Zitationsverhalten Zeiträume von mehreren (z.B. fünf) Jahren zugrunde gelegt werden [11].
- Bei der individuellen Leistungsbewertung sollten die „dos and don'ts“ der individuumsbezogenen Bibliometrie nach Glänzel & Wouters [12] beachtet werden.
- Die individuellen Beiträge jedes Autors auf Publikationen sollten in allen Zeitschriften standardisiert benannt

werden. Gefälligkeits- und Ehren-Autorenschaften sind unzulässig.

Ebene 1b

Mit jeweils individuell festzulegender Gewichtung sind als eigenständige Publikationsformen die folgenden Beiträge zu berücksichtigen:

- Monographien und Buchbeiträge
- Leitlinien und Health Technology Assessment-(HTA)-Berichte, auch wenn sie weder in Buchform noch in einer wissenschaftlichen Fachzeitschrift erschienen sind
- Publikationen von Originaldaten, Softwareentwicklungen, Patenten etc. (s. DORA)
- Verfügbarmachen wissenschaftlicher Erkenntnisse für die nicht-wissenschaftliche Praxis durch laiengerechte Aufbereitung; z.B. Patientenratgeber, Darstellungen in der Presse etc.

Ebene 1c

Die Zitation durch Leitlinien bei der unmittelbaren Begründung konkreter Empfehlungen („Leitlinien-Relevanz“) stellt ein geeignetes Maß des klinischen Impacts einer Publikation dar und sollte separate Berücksichtigung finden.

Ebene 2

Hier sind Aktivitäten zur intra- und interdisziplinären Vernetzung und Qualitätssicherung der Forschung zu bewerten.

Relevante Punkte sind hier insbesondere:

- Herausgabe wissenschaftlicher Zeitschriften als wesentliches Instrument zur Verbreitung von Forschungsergebnissen
- Mitarbeit in Vorständen, Sektionen und Arbeitsgemeinschaften von Fachgesellschaften und Wissenschaftsorganisationen
- Mitarbeit in wissenschaftlichen Beiräten anerkannter nationaler oder internationaler Wissenschafts- und Forschungsförderorganisationen
- Herausgehobene Gutachtertätigkeit (z.B. DFG-Kollegiaten)

Ebene 3:

Die Organisation und Leitung von Fachtagungen als wichtiges Medium der Wissenschaftskommunikation soll als eigenständige Leistung berücksichtigt werden.

3.3.2 Evaluation des „Inputs“

Inputbezogene Parameter für die Forschungsleistung können auf unterschiedlichen Ebenen definiert werden. Dabei sollen die vom zu Evaluierenden direkt beeinflussbaren Faktoren im Vordergrund stehen:

- Rahmenbedingungen (Grundausrüstung, Expertise, Strategiekonzepte, Anteil Arbeitszeit für Forschung, Betreuungsqualität für den wissenschaftlichen Nachwuchs); beeinflussbar auf Ebene von Fakultäten oder Zentren. Hierzu wird im vorliegenden Papier nicht detailliert Stellung genommen.
- Einwerbung und effektive Verwendung von Drittmitteln; beeinflussbar auf Ebene der einzelnen Forscher

Evaluationsparameter für die Forschungsleistung innerhalb der Fakultäten:

- Die Anerkennung und Gewichtung Evaluations-relevanter Drittmittel muss explizit festgelegt und transparent gehandhabt werden.
- Bei der Gewichtung Evaluations-relevanter Drittmittel müssen auf einer unabhängigen Begutachtung von Förderanträgen basierende Drittmittel höher bewertet werden als solche, die ohne unabhängige Begutachtung eingeworben wurden. Die Art des Drittmittelgebers sollte gewichtet werden: Öffentliche Förderung und neutrale Stiftungen sollten höher gewichtet werden als Interessenverbände oder industrielle Förderer.
- Auftragsforschung ist mit der Mittelbereitstellung ausreichend abgegolten und ist geeignet, Vorarbeiten für kompetitive Anträge zu finanzieren. Sie rechtfertigt keinen zusätzlichen Bonus aus öffentlichen Quellen, kann aber im Bereich der angewandten Forschung zur Bewertung von Personen herangezogen werden.
- Eine faire und transparente Bewertung der individuellen Mittelanteile an kooperativen Forschungsprojekten und industrie-unabhängigen Multicenterstudien soll sowohl die Leistung der Haupt-Antragsteller als auch diejenige der Kooperationspartner angemessen abbilden. Dies ist eine notwendige Voraussetzung für eine Kultur wissenschaftlicher Zusammenarbeit. Ggfs. sind allgemeine Gewichtungsschlüssel für die Beteiligung an unterschiedlichen Verbundtypen und Studien zu definieren (z.B. 1/3 der Bewertungspunkte für Hauptantragsteller und Verteilung der restlichen 2/3 auf alle Kooperationspartner bzw. Studienzentren).
- Je eingeworbener Wissenschaftlerstelle sollten gleiche Anteile berücksichtigt werden, da z.B. die Vorhaltung einer besonders teuren Technik kein wissenschaftliches Qualitätsmerkmal darstellt. Bei klinischen Studien kann eine Gewichtung nach Höhe der ausbezahlten Fallgelder erwogen werden, bei wissenschaftlichen Dienstleistungen können analog die Untersuchungskostenpauschalen eingesetzt werden.
- Die Kosteneffizienz der eingesetzten Mittel bezogen auf den wissenschaftlichen Output (gemäß den Kriterien unter „Impact“ und „Nachwuchsförderung“) sollte als Maß für den angemessenen Umgang mit bereitgestellten Ressourcen in die Leistungsbewertung von Forschern und Institutionen eingehen. Hierfür sind geeignete Algorithmen zu entwickeln.
- In der Evaluation medizinischer Forschung sollte ein Bonus für systemische, translationale und Probanden- oder Patienten-bezogene Forschung mit konkretem Bezug zur praktischen Medizin eingeführt werden.

3.3.3 Evaluation der Nachwuchsförderung

Wesentlicher Gegenstand der Evaluation ist die Nachwuchsförderung vom Anfängerstadium bis zum elaborierten Forscher.

Grundprinzip sollte sein, Nachwuchs bereits früh für die Wissenschaft zu begeistern und nachhaltig bis zur Ordinariatskompetenz zu fördern. Zielgruppen der Förderung sind Studierende, Doktoranden und Postdoktoranden sowie habilitierte Wissenschaftler der Human- und Zahnmedizin, (Zahn-)Ärzte und Angehörige benachbarter Gebiete. Dabei sollte ein mehrdimensionales Bild von Forschung vermittelt werden, das die wissenschaftliche Arbeit, die Anwendung in der praktischen Medizin, lebenslanges Lernen und Lehren umfasst. Dies deckt sich mit dem CanMEDs-Ausbildungsmodell [13], in dem die Funktion des „Scholar“ in Abbildung der internationalen Outcome-Frameworks als eine wesentliche Rolle der Absolventen medizinischer Studiengänge definiert wird. Insofern ist bereits die Förderung der wissenschaftlichen Kompetenz im Rahmen der curricularen Lehre im Medizinstudium als Instrument der Nachwuchsförderung zu betrachten. Zu ihrer Evaluation wird hier im Weiteren jedoch nur punktuell unter dem Aspekt der Forschungsrelevanz Stellung genommen, da das Thema der Leistungsevaluation in der Lehre Gegenstand einer eigenständigen, gemeinsam mit dem Medizinischen Fakultätentag in Vorbereitung befindlichen Stellungnahme sein wird.

3.3.3.1 Quantitative Indikatoren

Auf Institutionsebene (Fakultät, Institut, Klinik) erfassbare Merkmale der Nachwuchsförderung in der Forschung:

- Breite und Dichte von Maßnahmen zur Nachwuchsförderung, z.B.
 - Strukturierte Programme zur Heranführung an die Forschung: Austauschprogramme, curriculare (Vermittlung der Scholar-Kompetenzen, Pflicht- und Wahlveranstaltungen) und hypothesengestützte Angebote im Studium (u.a. Journal-Clubs, Haus-/Masterarbeiten, How-to Angebote), Graduiertenschulen/MD/PhD-Programme [14], Clinical/Physician Scientist Programme [15], interdisziplinär vernetzte Mentoringprogramme
 - Anzahl bereits in der studentischen Ausbildung begleiteter Studierender (als modulare Brücke vor Beginn wissenschaftlicher Profilierung [16])
 - Umfang gewährleisteter „Protected Time“ [17], also für Forschungsaufgaben faktisch reservierter Zeit ohne konkurrierende Verpflichtungen in Lehre oder Krankenversorgung, für Nachwuchswissenschaftler aller Karrierestufen bis zum Ordinariat. Indikatoren: Freistellungszeiten in Vollkraftanteilen pro Beschäftigungsdauer, Umfang einrichtungswweiter Forschungskonferenzzeiten pro Woche bzw. Monat mit prozentualer Angabe der Beteiligung der Nachwuchswissenschaftler

- Ergebnisse von Maßnahmen zur Nachwuchsförderung in der Forschung, z.B.
 - Zahl der Absolventen der o.g. Programme
 - Nachhaltigkeit der Programme; Indikatoren z.B. Karrierewege, Publikationen sowie Drittmittelwerbungen der Nachwuchswissenschaftler
 - Zahl der Tenure Track-Professuren in Forschung und Lehre [18]
 - Anzahl der Berufungen von Mitarbeitern auf wissenschaftliche Leitungspositionen bzw. auf klinische Leitungspositionen mit Mindest-Umsetzung der drei Dimensionen der Scholar-Kompetenz (Anwendung wissenschaftlicher Ergebnisse, lebenslanges Lernen, Lehren)

Individuell erfassbare Kriterien der Nachwuchsförderung

- Zahl adäquat betreuter Qualifikationsarbeiten (Indikator z.B. Promotionsverträge pro abgeschlossener Promotion)
- Zahl der eigenen Mitarbeiter mit Förderung in strukturierten Forschungs- bzw. Förderprogrammen bzw. mit Federführung in evidenzbasierten Leitlinienverfahren
- Karrierewege und Forschungserfolge der eigenen Mitarbeiter (Kriterien wie oben)

3.3.3.2 Qualitative Indikatoren

Auf Institutionsebene (Fakultät, Institut, Klinik) erfassbare Merkmale

- Vorhandensein und (horizontale wie vertikale) Vernetzung geeigneter Fördermaßnahmen für Nachwuchswissenschaftler auf den unterschiedlichen Karrierestufen, z.B. strukturierte Promotionsverfahren und nachhaltige Förderprogramme, interne und externe Peer Review- und „Coaching“-Verfahren („Lernen vom Besten“, „Common Sense“ zur Wissenschaftlichen Exzellenz, Entwicklung eigener „soft skills“) zum kollegialen Austausch auf Augenhöhe mit Nachwuchswissenschaftlern und Experten entsprechend des Leitfadens des Wissenschaftsrates und der Bundesärztekammer [19]; [20], Research Tracks, Vertrauensdozenten der Fachgesellschaften an der Institution.
- Vorhandensein und Zugänglichkeit von Qualitätsstandards (z.B. gute wissenschaftliche Praxis) und wissenschaftlicher Infrastruktur incl. Anschubfinanzierungsprogrammen (z.B. für 1 Jahr, um junge Wissenschaftler auf dem Weg zum DFG Antrag zu finanzieren), direkt antragsbezogener Infrastruktur (u.a. Kurse zur guten klinischen und wissenschaftlichen Praxis incl. guter Laborpraxis, Antragsberatung und Antragsunterstützung bei der Einreichung von Unterlagen bei Tierschutzbehörden, Verfügbarkeit elektronischer Laborbücher, Nutzung von u.a. Core Facilities/Core Research Units) und klinischer Studien-Infrastruktur (zur Unterstützung bei formaler Abarbeitung der Behördenarbeit: u.a. Datenschutz, Ethikkommission, Bundesinstitut für Arzneimittel und Medizinprodukte, Meldung bei den Landesbehörden, An- und Abmeldung bei Clinicaltrials.gov, Schreiben der Sicherheitsberichte; Verfahren

zur Vorveröffentlichung der Methodik, evidenzbasierte Prüfung der Unterlagen).

- Vorhandensein von Transparenz- und Gleichstellungsmaßnahmen, z.B. Transparenz des wissenschaftlichen Profils als Instrument für eine gezielte Hochschulauswahl durch Nachwuchskräfte, flache Hierarchien mit u.a. Nachwuchsvertretungen/wiss. „Assistentensprechern“, Tandemprofessuren (z.B. nach Schweizer Beispiel: Assistenzprofessur mit 50% Forschung und Lehre/50% Krankenversorgung, dann Tenure Track, 70% Forschung und Lehre und 30% Krankenversorgung und umgekehrt); Finanzielle Gleichstellung wissenschaftlicher und klinischer Tätigkeit im gesamten Karriereverlauf; Nachhaltigkeit durch Vereinbarkeit von Karriere und Familie (Kinderbetreuung, ggfs. auch nachts, an Wochenenden mit Priorität für WissenschaftlerInnen, Kinderbetreuung während Tagungen, Seminaren und Konferenzen, Kriseninterventions-Kinderbetreuung).

Individuell erfassbare Kriterien

- Aktive Mitwirkung an der Nachwuchsförderung durch Mitgestaltung u.a. von DFG-Nachwuchsakademien, Sommerschulen für herausragende Doktoranden, persönliches Engagement als Vorbildfunktion, Vermittlung grundlegender Wissenschaftskompetenzen an die Mitarbeiter, z.B. der Differenz zwischen Praxis (Alltagswissen), Profession (professionelles Wissen) und Wissenschaft (wissenschaftliches Wissen) sowie einer Brücke von Analytik, Transformation und Theorie, Frühzeitige Integration der Nachwuchswissenschaftler in Arbeitsgruppen mit zunehmender Verantwortlichkeit.
- Qualität der Forschungsleistung der Mitarbeiter, z.B. Reproduzierbarkeit der Ergebnisse der eigenen Nachwuchswissenschaftler durch andere Forschungsgruppen, Self-Assessment: Bewertung eigener wichtigster Forschungsergebnisse und Unabhängigkeit bei der eigenen Forschung durch Nachwuchswissenschaftler als Maß der Nachwuchsförderung durch den jeweiligen Mentor, Konsequenter Track der Forschung von Nachwuchswissenschaftlern, der sich in einer Forschungslinie widerspiegelt und die Relevanz der Forschung darstellbar macht.

3.4 Konsequenzen der Evaluation

Mögliche Konsequenzen der Evaluation sind a priori klar zu definieren und sollten sich an der Zielsetzung und Aussagekraft der Evaluation orientieren. Sie können neben dem unmittelbaren Feedback und gemeinsamer Diskussion der Ergebnisse im gezielten Einsatz von Instrumenten der Organisations- (z.B. Schwerpunktbildung), Projekt- (z.B. Förderentscheidung) und Karriereplanung (z.B. Berufung, Tenure, Mentoring) liegen, sofern diesem ein ausgewogenes und transparentes Verfahren zugrunde liegt.

Die leistungsorientierte Mittelvergabe ist nur eine von zahlreichen möglichen Konsequenzen der Evaluation und in ihrer Steuerungswirkung umstritten [21]. Besonders

wichtig für die in aller Regel intrinsisch hoch motivierten WissenschaftlerInnen ist die Wertschätzung der geleisteten Arbeit. Dabei sollten gleichermaßen Leistungen in der Forschung und der Lehre sowie (in der klinischen Medizin) auch Leistungen in der Krankenversorgung anerkannt werden. Die Anerkennung kann sich auch in der Bereitstellung zeitlicher Ressourcen darstellen. Dagegen läuft ein Übergewicht auf finanziellen Anreizen als extrinsischem Motivator langfristig Gefahr, intrinsische Motivation zu untergraben. Dies gilt insbesondere, wenn die zugrundeliegenden Bewertungsprozesse als intransparent oder ungerecht erlebt werden.

Es wird daher großer Wert auf eine ausreichende Grundausstattung der Einrichtungen gelegt, die jeweils nur in längeren Abständen auf der Basis umfassender Evaluationen an neue Entwicklungen angepasst werden sollte.

Anmerkungen

Danksagung

Das vorliegende Positionspapier basiert auf den Ergebnissen des Berliner Forums der AWMF zum Thema „Methoden zur Evaluation der medizinischen Forschungsleistung“ am 18.10.2013. Auf diesem Forum wurden von den eingeladenen Experten, VertreterInnen zentraler Wissenschaftsorganisationen (Deutsche Forschungsgemeinschaft, Medizinischer Fakultätentag, Projektträger im DLR, Wissenschaftsrat) und TeilnehmerInnen der ad-hoc-Schreibgruppen Vorschläge erarbeitet, die in anschließenden Diskussionen in der Kommission für Leistungsevaluation in Forschung und Lehre weiterentwickelt und vom AWMF-Präsidium konsentiert wurden. Die Autoren danken allen Mitwirkenden für ihre Beiträge zum vorliegenden Papier. Besondere Erwähnung verdienen (in alphabetischer Reihenfolge): Prof. K.-M. Debatin, Prof. R. Deinzer, Prof. W. Glänzel, Prof. C. Graf, Prof. H.-J. Heinze, Prof. S. Hornbostel, Dr. M. Kordel-Bödiger, Dr. T. Kostuj, Prof. H. K. Kroemer, Dr. A. Lücke, Prof. P. Meier-Abt, Dr. S. Moritz, Prof. A. F. J. van Raan, Prof. K. Rahn, Prof. G. Theilmeier, Dr. W. Warmuth.

Interessenkonflikte

Die Autoren erklären folgende Interessenkonflikte:

Alle Autoren sind Universitätsangehörige mit zugehörigen Ämtern in Leitung und Selbstverwaltung und Mitglieder wissenschaftlicher Fachgesellschaften, teilweise in leitender Funktion (THL, RDT, CFV). CHL, RDT und CS sind Mitglieder des AWMF-Präsidiums. Alle Autoren sind Autoren wissenschaftlicher Publikationen. Im Zuge der Erstellung dieses Positionspapiers haben CHL, SH, RDT und HJW von der AWMF bzw. einer Mitgliedsgesellschaft Reisekosten zu Sitzungen erhalten. CHL, RDT und CS haben Projektförderungen aus öffentlichen Mitteln erhalten. CHL, SH und RDT haben Fördermittel von Unternehmen bzw. Verbänden der privaten Wirtschaft erhalten. CHL

und RDT haben Honorare von Unternehmen der privaten Wirtschaft erhalten. CS war bis Mai 2014 Studiendekanin, RDT ist Prodekan für Forschung. CHL, EB, THL, RDT, CFV und CS sind Herausgeber bzw. Mitherausgeber wissenschaftlicher Zeitschriften. CHL und EB sind Buchherausgeber. CHL, EB, RDT, CFV und CS sind Gutachter für Forschungsförderorganisationen bzw. Stiftungen. THL ist Stiftungsbeirat.

Autorenschaft

Die Nennung der Coautoren erfolgt in alphabetischer Reihenfolge.

Literatur

- Frömter E, Brähler E, Langenbeck U, Meenen NM, Usadel KH. Das AWMF-Modell zur Evaluierung publizierter Forschungsbeiträge in der Medizin [The AWMF model for the evaluation of published research papers in medicine. Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften (Working Group of the Scientific Medical Specialty Societies)]. Dtsch Med Wochenschr. 1999 Jul;124(30):910-5.
- Brähler E, Strauss B. Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten : Eine aktuelle Übersicht [Performance-oriented allocations of financial resources at medical schools: an overview]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2009 Sep;52(9):910-6. DOI: 10.1007/s00103-009-0918-1
- Vahl CF. Forschungsbewertung: Fairness für forschende Chirurgen: ein Plädoyer. Dtsch Arztebl. 2008;105(12):A-625-8.
- Adler R, Ewing J, Taylor P. Citation Statistics – A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Stat Sci. 2009;24(1):1-14. DOI: 10.1214/09-STS285
- Alberts B. Impact factor distortions. Science. 2013 May;340(6134):787. DOI: 10.1126/science.1240319
- Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014 Jan;383(9913):267-76. DOI: 10.1016/S0140-6736(13)62228-X
- Wissenschaftsrat. Empfehlungen zur Zukunft des Forschungsratings. Drs. 3409-13, Mainz 25.10.2013. Available from: <http://www.wissenschaftsrat.de/download/archiv/3409-13.pdf>
- Van Raan AFJ. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. Technikfolgenabschätzung – Theorie und Praxis. 2003;12(1):20-9.
- Van Raan AFJ. Statistical Properties of Bibliometric Indicators: Research Group Indicator Distributions and Correlations. J Am Soc Inf Sci Technol. 2006;57(3):408-430. DOI: 10.1002/asi.20284
- Schulze N, Michels C, Frietsch R, Schmoch U, Conchi S. 2. Indikatorbericht Bibliometrische Indikatoren für den PFI Monitoring Bericht 2013. Hintergrundbericht für das Bundesministerium für Bildung und Forschung (BMBF). Berlin, Karlsruhe, Bielefeld: Institut für Forschungsinformation und Qualitätssicherung (iFQ), Fraunhofer-Institut für System- und Innovationsforschung ISI und Universität Bielefeld, Institut für Wissenschafts- und Technikforschung (IWT); 2012 [cited 23.6.2014]. Available from: http://www.bmbf.de/pubRD/Indikatorbericht_PFI_2013.pdf
- Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics. 2006;67(3):491-502. DOI: 10.1556/Scient.67.2006.3.10
- Glänzel W, Wouters P. The dos and don'ts in individual level bibliometrics. Paper presented at the 14th ISSI Conference, Vienna, 15-18 July 2013. Available from: <http://de.slideshare.net/paulwouters1/issi2013-wg-pw>
- Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007 Sep;29(7):642-7. DOI: 10.1080/01421590701746983
- Basu Ray I, Henry TL, Davis W, Alam J, Amedee RG, Pinsky WW. Consolidated academic and research exposition: a pilot study of an innovative education method to increase residents' research involvement. Ochsner J. 2012;12(4):367-72.
- Gaehtgens C. Clinical Scientist – Neue Karrierewege in der Hochschulmedizin. Werkstattgespräch am 27.-28.9.2013 in Schloss Herrenhausen, Hannover. Ergebnisse und Schlussfolgerungen. [cited 20.5.2014]. Available from: http://www.volkswagenstiftung.de/fileadmin/downloads/publikationen/veranstaltungsberichte/Veranstaltungsbericht_Clinical_Scientist.pdf
- Bierer SB, Chen HC. How to measure success: the impact of scholarly concentrations on students—a literature review. Acad Med. 2010 Mar;85(3):438-52. DOI: 10.1097/ACM.0b013e3181cccdbd4
- Sullivan R, Badwe RA, Rath GK, Pramesh CS, Shanta V, Digumarti R, D'Cruz A, Sharma SC, Viswanath L, Shet A, Vijayakumar M, Lewison G, Chandry M, Kulkarni P, Bardia MR, Kumar S, Sarin R, Sebastian P, Dhillon PK, Rajaraman P, Trimble EL, Aggarwal A, Vijaykumar DK, Purushotham AD. Cancer research in India: national priorities, global results. Lancet Oncol. 2014 May;15(6):e213-22. DOI: 10.1016/S1470-2045(14)70109-3
- Kelley WN, Stross JK. Faculty tracks and academic success. Ann Intern Med. 1992 Apr;116(8):654-9. DOI: 10.7326/0003-4819-116-8-654
- Wissenschaftsrat. Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Drs. 1656-11, Halle 11.11.2011. Available from: <http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf>
- Bundesärztekammer. Curriculum Ärztliches Peer Review. 2. Auflage. Berlin: Bundesärztekammer; 2013. (Texte und Materialien der Bundesärztekammer zur Fortbildung und Weiterbildung; 30). Available from: <http://www.bundesaerztekammer.de/downloads/CurrAerztlPeerReview2013.pdf>
- Krempkow R, Schulz P. Welche Effekte hat die leistungsorientierte Mittelvergabe? Das Beispiel der medizinischen Fakultäten Deutschlands. Hochschule. 2012;(2):122-42.

Korrespondenzadresse:

Prof. Dr. med. Christoph Herrmann-Lingen
Klinik für Psychosomatische Medizin und Psychotherapie,
Universitätsmedizin Göttingen, Von-Siebold-Str. 5, 37075
Göttingen, Deutschland, Tel.: +49-(0) 551-39-6707
cherrma@gwdg.de

Bitte zitieren als

Herrmann-Lingen C, Brunner E, Hildenbrand S, Loew TH, Raupach T, Spies C, Treede RD, Vahl CF, Wenz HJ. Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF). GMS Ger Med Sci. 2014;12:Doc11.
DOI: 10.3205/000196, URN: urn:nbn:de:0183-0001969

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/gms/2014-12/000196.shtml>

Eingereicht: 23.06.2014

Veröffentlicht: 26.06.2014

Copyright

©2014 Herrmann-Lingen et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.