Eur Arch Psychiatry Clin Neurosci (2010) 260:25–39
DOI 10.1007/s00406-009-0070-9

SPECIAL ISSUE

Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations

Hans-Jürgen Möller · Wolfgang Maier

Published online: 17 October 2009

Abstract Psychopharmacotherapy should now be regulated in the sense of evidence-based medicine, as is the case in other areas of clinical treatment in medicine. In general this is a meaningful development, which principally will have a positive impact on routine health care in psychiatry. But several related problems should not be ignored. So far consensus on an internationally accepted evidence graduation could not be reached due to several difficulties related to this. For example, focussing on the results of meta-analyses instead of considering relevant single studies results in a decision-making logic which is in conflict with the rationale applied by drug authorities in the licensing process. Another example is the relevance of placebo-controlled trials: if randomized placebo-controlled phase-III studies are prioritized in the evidence grading, the evidence possibly deviates too far from the conditions of routine clinical care due to the special selection of patients in those studies. However, a grading primarily based on active comparator trials could lead to wrong conclusions about efficacy. This concerns especially the so-called "effectiveness" studies and other forms of phase-IV studies with their less restrictive methodological rigidity. Attempts to regulate psychopharmacotherapy in the sense of evidence-based medicine come closer to their limits the more complex the clinical situation and the respective decision-making logic are. Even in times of evidence-based medicine a large part of complex clinical decision-making in psychopharmacotherapy still relies more on clinical experience and a consensus on clinical experience, traditions and belief systems than on results of efficacy oriented phase-III and effectiveness-oriented phase-IV clinical studies.

Keywords Evidence-based · Evidence grading · Psychopharmacotherapy · Guidelines

H.-J. Möller (✉)
Department of Psychiatry, Ludwig-Maximilians-University, Nussbaumstrasse 7, 80336 Munich, Germany
e-mail: hans-juergen.moeller@med.uni-muenchen.de

W. Maier
Department of Psychiatry, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany
e-mail: Wolfgang.Maier@ukb.uni-bonn.de

Introduction

'Evidence-based medicine' (EBM) has become one of the most important terms in today's health system [139] and thus also in psychiatry. In this context, 'evidence' is understood as the sum of empirical knowledge available on a certain issue derived from adequately designed clinical trials. This definition is in contrast to the common sense terminology in which the term 'evidence' has the meaning: it is evidence for me based on what I have seen. Especially clinicians use the term traditionally as a synonym for their personal clinical observations and experiences or the clinical experiences of a group of clinicians.

Given the increasing importance of EBM, it seems meaningful to analyse possibilities, problems and limitations of this approach. Of course, due to limitations of space, only some of the most relevant aspects can be covered.

Central elements and basic problems of evidence-based medicine in psychopharmacotherapy

Evidence in the sense of EBM is the result of a critical and systematic overall evaluation ('critical appraisal') of

(published) results of scientific studies. The evidence may refer to various areas such as diagnosis or treatment. The complex medical knowledge about a certain area is evaluated and summarized in meta-analyses and systematic reviews, etc., and then transposed by specialist groups into recommendations and guidelines [6, 26, 27, 57, 80, 86, 106]. The physician should base his diagnostic and therapeutic decisions on these evidence-based recommendations, giving an empirical foundation to his actions and making them more rational [35, 60]. In this way, the scope for subjective judgment primarily based on so-called clinical experience or other belief systems should be reduced as far as possible.

In the context of EBM, randomized control-group studies are considered to be the decisive level of scientifically proven evidence as far as therapeutic aspects are concerned [90]. The knowledge gained from non-interventional (observational) studies as well as from single-cases studies is only seen as relevant as an addition to such studies or as a replacement in cases where empirical studies of a higher methodological degree are lacking. This view corresponds to the general methodological understanding of empirical research. Evidence graduation is geared to the fact that for methodological reasons certain study designs yield results that are more likely to be reliable. This corresponds with the rules of the methodology of empirical research [24, 42]. Thus randomized, control-group studies have a higher value than non-randomized or uncontrolled studies.

Evidence-based medicine and the associated treatment recommendations and guidelines as well as other concurrent summaries of the available knowledge have become an important part of quality improvement and quality assurance in psychiatry, particularly in the area of psychopharmacotherapy. This guarantees an adequate rationality of treatment [60]. There is no doubt that this is an important subject and that guidelines can be very helpful for a physician's decision-making in face of the complex and complicated amount of knowledge about the treatment of a certain psychiatric disorder. However, it was questioned whether the current EBM movement is going too far and underestimates the value of clinical experience in solving practically relevant issues [68]. Even Sackett, one of the fathers of EBM, stressed the point that EBM means "integrating individual clinical expertise with the best available external clinical evidence from systematic research" [140].

Over recent years, many national and international psychiatric associations have published treatment recommendations or guidelines. The practice guidelines of the American Psychiatric Association (APA), which are used far beyond American psychiatry, are an example of such guidelines. The guidelines of the World Federation of Societies of Biological Psychiatry (WFSBP) [15, 16, 18–20, 44, 45, 64–67] are also gaining increasing international recognition. Although these and several other guidelines [51] follow an EBM approach, they often are not consistent concerning the evidence grading and the respective treatment suggestions, which underlines the fact that besides the empirical database, interpretations of the huge data sets play an important role.

An important aspect in the preparation of guidelines is the establishment of standards to be adhered to during the development of guidelines. Publications concerning the development of the scientific content of guidelines—how to extract the literature, how to summarize the literature, how to come to valid conclusions, etc.,—are highly demanded, and several suggestions have been made. Based on this, a quality ranking of guidelines can be performed [51].

Organisational-technical factors of guideline development such as the procedure for selecting experts for the respective expert commissions and the way in which a consensus is reached in these bodies as well as the possibility that they could be influenced from various sides are of great relevance and can have a significant influence on the result. The 'evidence' is often a biased opinion on the available data from the standpoint of the various stakeholders.

Attempts to influence these rather organizational-technical factors can be made by various groups. The pharmaceutical industry [39, 61, 77, 88, 112, 130, 132, 153, 157], time and again criticised for trying to influence physicians' prescribing practices, is not the only potential faction. Governmental health institutions, physicians' associations, insurance companies, health care trusts and others are often not as 'neutral' [116, 119] in this respect as they claim to be, which can easily be proven, e.g., by the design and outcome of some effectiveness studies on antipsychotics funded by public sponsors [117, 120]. As another example, one only has to think of the ongoing debate about the potential advantages of second-generation antipsychotics and the contradictory positions of various groups [118, 149].

Another relevant problem is that guidelines have a retrospective perception of the situation, particularly when they take a longer time to develop. They therefore tend to make conservative decisions about treatment and do not adequately accommodate new advances. This is of particular practical relevance whenever increasing requirements are placed on the guideline development process, so that it can take 2–3 years to develop and finalize a guideline. Because the experts involved cannot continuously afford to spend a lot of time on such guideline projects, and the costs of developing guidelines are enormous, they are often not revised for several years. In view of the known short half-life of medical knowledge, such a long period of validity for guidelines cannot be justified. Guidelines have to be revised at least after 3–4 years.

It is also important to disseminate the guidelines among the relevant group of physicians and to encourage physicians to follow them through specific implementation strategies. It is known that the transfer into clinical practice is associated with many problems [62, 63]. For example, the guideline from the American Diabetes Association and the APA on the problem of weight gain and associated metabolic syndrome during treatment with neuroleptics [4] did not result in the suggested checks being performed by American psychiatrists [37]. To increase the chance of acceptance, guidelines should be written in a user-friendly way [13].

Quality assurance procedures have to be implemented in health systems parallel to the development of EBM-based guidelines in order to put the EBM approach into reality [129]. Self-control procedures by the medical fraternity, but perhaps in future also external control by respective health system institutions, should guarantee that when making decisions physicians follow EBM as defined in treatment recommendations and guidelines. However, although this approach appears to make sense it will be scrutinized critically beforehand because of the normative implications, taking away from the doctors a large part of their freedom in clinical decision-making. It was even questioned whether guidelines in general improve the quality and outcome of treatment, and the demand was made that each guideline has to prove this before being widely introduced [101].

Even being open to such an EBM development, it has to be questioned whether the existing guidelines are on such a quality level that it is meaningful to force doctors to follow them. Can 'evidence' predominantly derived from phase III studies with all their well-known limitations in terms of generalisability really be seen as the basis for everyday clinical decision-making on patient populations which are, concerning different characteristics, quite unsimilar to the phase III study samples? In this regard, the guidelines should possibly better define the kind of patient population for which they are valid. Among other critically discussed practical issues, the financial implications are of great importance for the prescribing physician. The guideline might prioritise for good reasons (better efficacy and/or better tolerability) a new drug, but the doctor cannot prescribe this expensive drug and has to replace it with an older/cheaper drug for economic reasons.

Even if treatment recommendations and guidelines, as opposed to directives, do not oblige the physician to use the prioritized treatment option, there is some justifiable concern that they could be over-interpreted by socially relevant bodies in the health system and thus have further-reaching consequences—for example, that a health insurance company refuses to pay a certain therapeutic approach or that a physicians' association recommends that certain drugs should no longer be prescribed. To avoid too extreme consequences, guidelines are often written in a very diplomatic phraseology, which however undermines the clarity of the instructions.

Meta-analyses and systematic reviews as approaches to determine evidence and their limitations

The two most important approaches to determining evidence are systematic reviews and meta-analyses. Systematic reviews present in a narrative form a critical overview and qualitative evaluation of the studies available on a certain topic. The advantages and disadvantages of individual studies are weighed against each other, and an overall opinion deduced on the basis of a qualitative description of the various study results is presented: X has better efficacy than placebo or the same efficacy as another drug. In the context of EBM, systematic reviews have to fulfil high methodological demands regarding both the consideration of all relevant studies and their critical evaluation. The demands placed on such reviews therefore go beyond those placed on traditional reviews on a topic.

Meta-analyses combine quantitatively the results of studies on a specific subject of interest judged to be methodologically adequate for inclusion. This can be related to all fields of interest, mostly clinical outcomes, results of drug trials, but also, e.g., to neurobiological parameters [38, 145]. They calculate an 'effect size' or other quantitative measures, which show, e.g., the quantitative difference between two comparative treatment approaches (i.e., vs. placebo or vs. standard drug). The comparison of effect sizes assumes that they have been taken from the same basic population. This requirement is mostly only roughly fulfilled, since the various studies being combined have different designs and mostly different basic conditions, e.g., with respect to setting variables, inclusion and exclusion criteria, pre-treatment, and concomitant medication.

The results of meta-analyses are becoming increasingly important for the development of guidelines and textbooks on the basis of EBM [156]. This may be because the quantitative summarizing of results into effect sizes is easier to convey than the differentiating, qualitative conclusions formed on the basis of systematic reviews. Compared to narrative systematic reviews, meta-analyses really do have the advantage that they can condense the results to quantitative core values (effect sizes) while reviews can only draw qualitative conclusions. Even so, meta-analyses cannot replace narrative systematic reviews, which have the advantage of being able to consider in a differentiated way the special conditions of the individual studies with respect to study design, patient selection and drug dosing,

etc. However, this detailed analysis requires expert clinical-psychopharmacological knowledge and a highly detailed description, including all kinds of confounders (e.g., different prescription pattern of rescue medication in the different treatment arms) which are not always detectable in the relatively short systematic reviews which sometimes precede meta-analyses. Both procedures should be regarded complementarily. A leading position of meta-analyses compared to the more extensive narrative systematic reviews is not justifiable [48, 104, 105, 112, 144].

Meta-analyses can lead to inconsistent results, solely depending on how the data are collected and on the different statistical methods of meta-analysis [104]. In addition, the choice of specific selection criteria for inclusion/exclusion can result in a "cherry picking" unintentionally or even intentionally. Meta-analyses do not respect whether an outcome criterium was defined as being primary or secondary in the original study, which is a misunderstanding of the experimental approach leading to severe consequences. Often the comedication, which can play an important role, e.g., especially in the placebo group, thus confounding the homogeneity of the compared groups, is more or less neglected [118]. Because of the inclusion of irrelevant or methodologically problematic studies, the noise can be increased in such a way that the detection of a meaningful signal is extremely reduced. An example is the inclusion of so-called failed studies in meta-analyses, e.g. [87, 153]. Failed studies are studies in which neither the experimental drug nor the standard comparator reach superiority to placebo, because these studies are apparently performed in a drug-insensitive sample. On the other hand, meta-analyses can also lead to false positive results by artificially inflating the statistical power. For example, in a meta-analysis on placebo-controlled studies on paroxetine with a positive efficacy result only one of the three included original studies had positive results in the LOCF analysis [141]. In another meta-analysis on placebo-controlled studies on lamotrigine in acute bipolar depression a positive efficacy result was obtained [54], although four original studies were negative and one was positive only in the secondary efficacy parameter. Meta-analyses also do not take into consideration the stage of development of the application/evaluation of a drug: often in the early stage a drug is dosed too low or too high, which can have severe consequences in terms of efficacy or side effects (e.g., risperidone, quetiapine). In general, dose aspects are difficult to consider in meta-analyses, because they are often not dose-linear, but follow *u*-curves, for example. Sensitivity analyses to control for dose effects or other influencing factors/confounders are not sufficient and often methodologically not adequate.

Principally, a meta-analysis should be regarded like a clinical trial, which needs replication by at least another one, and not as the final answer. Experience has shown that often meta-analyses on the same topic and using more or less the same database come to contradictory results. This can be explained by different methodological aspects, among others factors related to the statistical methodology of meta-analyses. An example for this statement is the meta-analyses on the potential suicidality-increasing effects of antidepressants [124]. Also the meta-analytical results on the question of whether, for example, so-called dual antidepressants are more efficacious [12, 21, 30, 53, 72, 125] than selective serotonin receptor inhibitors or whether second-generation antipsychotics have advantages in terms of efficacy or extrapyramidal tolerability are largely inconsistent [38, 52, 53, 94, 95, 97, 98].

The numerical value of the effect size seems definite and significant but is in fact full of ambiguity, resulting among other things from basic methodical problems of meta-analyses. The seemingly handy and pictorial value of the effect size can too easily be interpreted in a naïve and simplifying or deliberately tendentious way because the immense and complex amount of clinical data behind it is no longer visible. Over-interpretation of effect size values or numbers needed to treat (NNTs) is often seen but is inappropriate because of various basic problems of meta-analyses and should therefore be questioned critically [99]. For example, differences in the effect sizes can only be related to the application of different scales [81, 120]. Especially if effect sizes or NNTs are only based on one study with all its specific conditions [3, 31, 58] and not on a number of studies [143], they should not be seen as generalisable. Considering the heterogeneity of treatment effects [146], it is generally extremely problematic to make conclusions for an individual patient based on one huge study or a meta-analysis of several studies. The NNT approach can induce a false impression of accuracy. This problem of pseudo-accuracy should always be taken into consideration.

Treatment recommendations and guidelines are based on systematic reviews and meta-analyses of empirical knowledge and the related expert consensus. They show evaluations of the respective amount of evidence of empirical knowledge on a coincidence grading scale. It needs to be emphasized that evidence grading, and in particular grading of recommendations, are not trivial processes in which empirical data levels are practically transcribed one to one. They are rather complex processes, some of which go far beyond the level of data [9–11, 69–71, 103]. This applies to evidence grading and even more so to recommendation grading.

The evidence grading of empirical knowledge is in many guidelines connected to a recommendation for clinical action, which can be graded according to the degree to which the recommendation seems empirically justified but

also including other aspects like tolerability, practicability, etc. While already evidence grading has only a certain relationship to empirical data, recommendation grading surpasses these and incorporates personal treatment stereotypes, regional/national treatment traditions and other aspects, depending on the body designing the guidelines. In this way, recommendations can differ considerably from evidence of empirical data as regards content and grading. The problems associated with grading recommendations have resulted in some guidelines abstaining from specifying levels of recommendation, e.g., the NICE schizophrenia guidelines [131].

The problem of differences in evidence criteria and evidence grading

The set of criteria for the various evidence levels, which appear to be clearly formulated, are actually full of inconsistencies and are far from fulfilling an operational definition [16]. This becomes clear when one focuses on the respective details, which cannot be done here for reasons of space. The principal problem is that there is no uniform, internationally accepted definition of evidence and the associated evidence levels, even though the expression "evidence level I, II, III or A, B, C", etc., suggests that this is the case. Thus, the choice of evidence criteria and levels alone can yield very different results for the case at hand, as shown by a random selection of a few concrete examples (Tables 1, 2, 3).

EBM overall and many guidelines prefer to base their evidence on randomized controlled trials (RCTs). However, it often remains unclear whether results of placebo-controlled studies are seen as more important than those of non-placebo-controlled studies, although this would make sense in view of the rules of empirical research in psychopharmacotherapy and would correspond with the demands of the regulatory authorities. In addition, often the criterion of testing under double-blind conditions is not considered as a differentiating criterion in EBM-related publications (see below). A particularly controversial point is whether the results of important individual studies with outstanding methodology are superior to those of meta-analyses [32]. Most guidelines prefer results of meta-analyses alone or together with results from individual studies. Narrative, systematic reviews do not appear to play a significant role in evidence grading, or at least they are not mentioned in most of the evidence levels, although they would give important complementary views to the findings of meta-analyses, because they go more into the specific details of the original studies.

Table 1 NICE grading scheme and hierarchy of evidence [126]

Evidence category	Source
I	Evidence from: meta-analysis of randomised controlled trials, or at least one randomised controlled trial
II	Evidence from: at least one controlled study without randomisation, or at least one other type of quasi-experimental study
III	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case–control studies
IV	Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

Adapted from [41]

Results of meta-analyses versus results of single studies in the definition of the highest level of evidence

Many guidelines define the highest level of evidence using meta-analyses of randomized, controlled studies (Table 1). However, the prioritization of meta-analyses is not as unproblematic as it first appears [99, 104, 105, 114]. All the methodological problems discussed above should be considered in this context.

The basic prerequisite for a meta-analysis, and also for the subsequent utilization of the results, is the methodological stringency applied when it is performed. This applies to the systematic search for the studies to be included, the methodological evaluation of the identified studies and to the clinical evaluation of any heterogeneity that may appear, etc. For reasons of principle, meta-analyses that only include small, randomized, controlled studies or randomized studies of inferior methodological quality should be considered very cautiously. As already mentioned, a meta-analysis should be regarded like a clinical trial which needs replication by other meta-analyses, and not as a final answer. Given the extreme variation of results of meta-analyses on the same issue, it is not acceptable to stick only to the results of one meta-analysis.

In general, it is of great importance that results from meta-analyses on psychotropic drugs are discussed in a very careful way and with the knowledge and expertise of clinical psychiatrists/psychopharmacologists. The publication of a recent meta-analysis on the efficacy of antidepressants [87] underlines how misleading the interpretations of such meta-analytical results can be, if the authors do not have adequate experience and expertise [118]. A more careful and expert-like interpretation concerning the placebo-verum differences of antidepressants or of second-generation antipsychotics, respectively was presented in the meta-analyses by Melander et al. [113] and Leucht et al. [96, 98].

Table 2 Evidence criteria used in the WFSBP treatment guidelines, Part I [18]

Level A: This level is achieved if research-based evidence for efficacy is given from at least three moderately large, positive, randomized controlled (double-blind) studies (RCT). In addition, at least one of these three studies must be a well-conducted, placebo-controlled study
Level B: This includes evidence of efficacy from at least two moderately large randomized, double-blind studies (this can be either ≥two comparator studies or one comparator-controlled and one placebo-controlled study) or from one moderately large randomized, double-blind study (placebo-controlled or comparator-controlled) and ≥one prospective, moderately large (sample size of ≥50 participants), open-label, naturalistic study
Level C: This level is achieved if one randomized, double-blind study with a comparator treatment and one prospective, open-label study/case series (with a sample size of ≥10 participants) showed efficacy, or at least two prospective, open-label study/case series (with a sample size of ≥10 participants) showed efficacy
Level D: Expert opinion-based supported by at least one prospective, open-label study/case series (sample size ≥10 participants)
No level of evidence: Expert opinion for general treatment procedures and principles

Table 3 Evidence criteria used in the APA treatment guidelines [5]

[A] *Randomized, double-blind clinical trial*. A study of an intervention in which subjects are prospectively followed over time; there are treatment and control groups; subjects are randomly assigned to the two groups; and both the subjects and the investigators are "blind" to the assignments
[A−] *Randomized clinical trial*. Same as above but not double blind
[B] *Clinical trial*. A prospective study in which an intervention is made and the results of that intervention are tracked longitudinally. Does not meet standards for a randomized clinical trial
[C] *Cohort or longitudinal study*. A study in which subjects are prospectively followed over time without any specific intervention
[D] *Control study*. A study in which a group of patients and a group of control subjects are identified in the present and information about them is pursued retrospectively or backward in time
[E] *Review with secondary data analysis*. A structured analytic review of existing data, e.g. a meta-analysis or a decision analysis
[F] *Review*. A qualitative review and discussion of previously published literature without a quantitative synthesis of the data
[G] *Other*. Opinion-like essays, case reports, and other reports not categorized above

Of relevance in this context is the fact that the large regulatory authorities such as the American FDA and European EMEA do not recognize meta-analyses as the primary basis for deciding whether or not to license a drug. This is because of basic methodological/statistical considerations on the confirmative testing of hypotheses. The authorities base their decision on the results of single studies of sound methodology (particularly phase III studies, mostly placebo-controlled). The resulting conflicts are foreseeable: as an extreme example, a licensed substance could be classified as inefficacious when results of meta-analyses are applied in the context of EBM because this approach considers not only pivotal, phase III studies, as is the case in the licensing process, but also other studies that may have had different objectives and were often not performed with the primary objective of proving efficacy. In other words, an evidence-based guideline with treatment recommendations/guidelines which rates the results of meta-analyses as the highest evidence criterion may reach conclusions that differ from or even contradict those of the regulatory authorities since it follows a different approach to decision-making. An example of this is the situation of lamotrigine in the treatment of acute bipolar depression. Lamotrigine was licensed for relapse prevention of bipolar disorder based on sufficient data [22], but not for treatment of acute bipolar depression because most placebo-controlled studies were negative; however, a positive result in the meta-analysis by Geddes et al. [22, 54] could lead to a positive evidence grading, even in the first category.

The treatment guidelines of the World Federation of Biological Psychiatry, e.g., [18, 19, 44, 64] apply a different system of evidence grading (Table 2), which was originally used in the "Schizophrenia Patient Outcome Research Team" (PORT) treatment recommendations [91] and recently modified [15, 16]. The decisive difference to the evidence criteria of many other guidelines is that the highest evidence level is not based on the results of meta-analyses but on the results of important and methodologically sound single studies. Thus, the catalogue of evidence criteria corresponds in principle to that of the regulatory authorities. The APA Practice Guidelines also rate the evidence on the basis of results of single studies (Table 3).

It is extremely important that an internationally uniform evidence grading be established, although this will not be easy in view of the many problems described above. International working groups of methodologists, e.g. GRADE, are currently working on a standardization of evidence levels [9, 71]. This is a very complex and sophisticated approach.

Placebo-controlled studies versus standard comparative studies as a requirement for the highest level of evidence

In some guidelines, for example those of the WFSBP, double-blind placebo-controlled studies, besides other double-blind randomised control group studies, are a prerequisite for the highest level of evidence (Table 2). Other guidelines, for example the APA Practice Guidelines [5], require randomized control group studies (in particular randomized, control group studies comparing a new substance with a standard drug) (Table 3), without specifically demanding placebo-controlled studies, and some do not even demand explicitly double-blind conditions. The APA Practice Guidelines even differentiate only minimally between evidence from randomised, double-blind, control group studies, which are required for evidence level [A], and evidence from unblinded control group studies, which are required for evidence level [A−]. If placebo-controlled studies are grouped together in the highest evidence level with studies in which, e.g., a new substance is compared to a standard drug, without differentiating between blinded and unblinded studies, then study types of different validity are being put on a par with each other. This is not a meaningful approach since, at least in several psychiatric indications, e.g., depression, studies without a placebo control do not allow valid conclusions (internal validity) to be drawn; for this reason large regulatory authorities (e.g., FDA, EMEA) require placebo-controlled studies [14, 49].

On the other hand, a too one-sided and extensive overemphasis of the relevance of placebo-controlled studies is not desirable. Although they are often necessary for proof of efficacy in many indications, the generalisability of such study results to routine clinical care is often not guaranteed (the problem of internal vs. external validity). It is known that the placebo-controlled studies performed for the registration of new drugs in psychiatry are problematic in that they are not very closely related to routine clinical care and must therefore rather be seen as 'proof of concept' studies. Generally, in psychiatry only about 10% of patients who would be suitable for a study can be recruited, even in non-placebo-controlled, randomised, controlled, phase III studies [23, 25, 100, 133]. The recruited patients are also known to be selected in various ways (e.g., exclusion of comorbidity, of older patients, of patients at higher risk). The participating patients are therefore not even representative of the sample of patients of the respective diagnosis group at the study center who could principally be recruited, let alone of the whole group of all patients of this diagnosis group. The problem of selection increases with the intricacy of the study procedures. A placebo control is highly demanding and results in an especially high degree of selection. This is illustrated by the example of placebo-controlled studies in manic patients in which normally only patients with relatively mild manic symptoms are included. The same is true for placebo-controlled studies in depression, where patients with extremely severe depressive symptoms and/or suicidality are mostly not included, for example.

In order to avoid guidelines completely losing their relationship to clinical reality by preferring study types with too little generalisability, greater emphasis should be placed on other empirical research approaches. A drug that has been evaluated in placebo-controlled studies with the selection problems described above should also be tested in studies with less restrictive methodology, e.g., randomized, control group studies versus a standard drug; the results should at least show a tendency towards consistency. The three-arm study design recommended by the European regulatory authority, EMEA/CPMP [33], in which the experimental substance is compared to placebo and a standard drug, delivers more meaningful results but cannot avoid the problems associated with the extensive selection of patients since it still has a placebo group.

It should be remembered that, traditionally, there was a demand for a psychopharmaceutical drug to be clinically evaluated in a phase model at various methodological levels of empirical research and with approaches of different methodological stringency. This means that evidence for efficacy and tolerability should be obtained from phase IV studies, which are more closely oriented towards routine clinical care [7, 8, 40, 73, 74, 123], to complement the results of phase III studies with their strict methodology. In such a phase model of clinical/pharmacological evaluation, the evidence from each phase is seen to be complementary and part of the overall evidence. This idea can no longer be found in the systems currently used in guidelines to assess evidence, since evidence is rated according to the study design with the most demanding methodology for the respective therapy (e.g., placebo-controlled studies) without ascertaining whether consistent results are available from less restrictive but more generalisable study types. A grading of evidence that is more relevant for clinical reality should assess whether results are available from studies with both high internal (e.g., control group studies) and high external (e.g., observational studies) validity and whether the results are principally the same. So far, the current interest in effectiveness studies is principally positive [100, 136, 142]. However, the results of these

effectiveness studies should not be overinterpreted due to their principal methodological limitations (as demonstrated, e.g., for the CATIE trial, [117]. From a statistical point of view it also does not seem unproblematic that, e.g., the STAR*D study was used to publish about 100 publications answering different questions.

Differences in the rating of evidence in psychopharmacotherapy and psychotherapy

There is no room here to discuss the basic problems of efficacy research in psychiatry but only problems that arise in clinical psychopharmacology when effect sizes or evidence ratings are compared [17, 55, 75, 89, 92, 102, 154].

Now that it is becoming more common to calculate effect sizes to describe the empirical evaluation of psychotherapy/psychosocial therapy and to rate evidence levels [59], it is principally possible to compare these to the evidence criteria used in psychopharmacotherapy. But this carries the danger of absurd comparisons of effect sizes or evidence levels obtained from different methods of therapy evaluation. The effect sizes on the efficacy of certain psychotherapeutic treatments are derived from unblinded parallel control groups, if at all. Many studies only use the waiting list control group approach. Others do not use a control group design at all. It is obvious that the effect sizes derived from such studies cannot for principle reasons be compared to the effect sizes from placebo-controlled studies because the placebo approach minimizes the effect size.

If an evidence-grading system is used which does not demand placebo-controlled or double-blind control group studies as a criterium for the highest level of evidence, it could be that a certain psychotherapeutic method A would be assigned as the highest level of evidence for the treatment of depression. This would be misleading. In most guidelines on psychopharmacological treatment, an antidepressant can only achieve the highest evidence level in most evidence-grading systems on the basis of double-blind, randomized and, in some guidelines, placebo-controlled studies, whereas psychotherapeutic procedures are not evaluated under placebo-controlled or double-blind conditions. It is impossible for psychotherapy to achieve a rigid methodological approach of this level, for understandable reasons. This must have consequences for the attributed level of evidence, because due to the lack of placebo and double-blind conditions, different kinds of bias are possible. The different methodological basis of psychotherapy evaluation compared to the evaluation of psychotropic drugs implies that a direct comparison is impossible. In order to avoid confusion—for example, by evidence-grading systems specifically adapted to the situation of psychotherapy evaluation—it would be better to develop a standardized system for rating evidence for all treatment procedures in psychiatry in which psychotherapeutic approaches could not achieve the highest level per se because of the principal differences in their methodology (it is not possible to use a real placebo control and impossible to perform a double-blind study).

This is even more so for psychosocial procedures, which cannot even fulfill the requirement of randomized, control group studies because of their special characteristics and which use methodologically less restrictive evaluation procedures. Particularly in the area of research in psychiatric care there is only a limited amount of evidence even below this threshold [83]. It ranges from a lack of studies on supported housing [29], through three studies on day centers [28], nine studies on acute psychiatric day clinic treatment [109] and ten studies on intensive (or clinical) case management [108], up to 18 studies on supported employment [36] and 20 studies on assertive community treatment [107].

Can the complexity of clinical decision-making processes be evidence based?

The complexity of clinical decision processes goes far beyond what is actually evidence based. For this reason, in many areas treatment guidelines, which aim to come close to the complexity of clinical decision-making processes, have to replace the evidence-oriented approach with a consensus-oriented approach.

Efficacy and tolerability aspects when comparing different drugs

EBM already has much greater difficulties answering the very important question of the comparative efficacy or tolerability of two or more substances than the question of whether a substance has sufficient efficacy. The former is more complex and difficult to answer because hardly any multi-arm studies are performed to compare various substances directly. At the most, these are usually three-arm studies comparing the new drug versus a standard drug and placebo. For this reason, further-reaching conclusions that also include other substances can often only be drawn indirectly on the basis of additional comparative studies. For example, the new substance X was compared to placebo and standard substance A in a three-arm study and to substance B in a different study. If a multitude of different studies are performed, the database is large enough to allow direct and indirect comparisons with meta-analyses and thus to enable conclusions to be drawn about the relative efficacy and tolerability of the substances. For a drug which was not directly compared with a comparator of interest, an estimation is based on trials in which the drug

was compared to other drugs for which on the one hand comparator data with drug X are available, and on the other hand also comparative data with the comparator of interest exist. Such a multiple-treatment meta-analysis was recently published [30] for antidepressants focussing on efficacy and acceptability (measured by the dropout ratio). However, great care should be taken when interpreting indirect conclusions since many confounding factors can influence the results of studies. For example, selection problems can make it impossible to demonstrate the better efficacy of a new neuroleptic in treating negative symptoms if the average level of negative symptoms in the study sample was too small to allow any effects at all to be detectable.

Difficulties associated with comparative risk/benefit analysis

Complex clinical decision-making processes are often not only concerned with differences in efficacy, or the evidence levels with which the efficacy was evaluated, but also with tolerability aspects. The following example of the drug treatment of acute bipolar depression [56] makes clear how difficult it is to base more complex clinical/therapeutic decision-making processes on an empirical foundation in the sense of EBM. American psychiatry in particular advanced the position that patients with acute bipolar depression should not normally be treated with antidepressants because of the danger of switching into mania but should instead be treated with mood stabilizers. This recommendation was given in various guidelines even though the antidepressant efficacy of mood stabilizers was far from having been proven in the sense of the usual evaluations of an antidepressant and was thus not evidence based. Thus, tolerability aspects alone (switch risk) gave rise to a recommendation that may have deprived many patients of an efficacious treatment for their depression [42, 121, 122] who would not have had an increased switch risk if they had been given an SSRI [56]. This example shows that when guidelines attempt to deal with complex treatment decision-making processes they are often no longer sufficiently evidence based. Furthermore, one-sided evaluation processes based on certain school-related or national belief systems may sometimes play a more decisive role than the data itself.

Founding treatment algorithms on evidence

Attempts have been made to include more complicated treatment approaches, as are common in everyday psychiatry (e.g., comedication or sequential treatment approaches), in an evidence-based approach. The problems associated with the comparative evaluation of the efficacy and tolerability of individual drugs are even more relevant in the case of complex treatment procedures. For example, even the question of whether a comedication increases efficacy is not well investigated [34]. There are normally insufficient empirical data available to provide empirical proof of more complex treatment procedures. Similarly, the common approach of switching schizophrenic patients to a different neuroleptic with a different chemical structure or mechanism of action if their original neuroleptic shows insufficient efficacy has not yet undergone adequate empirical testing and at least not sufficient proof [34, 43, 84, 85, 115, 146–148]. The situation is similar for a switch from one antidepressant to another with a different mechanism of action: not enough data are available to make evidence-based decisions [78, 134, 137, 155]. The situation concerning combination/augmentation therapies is similar.

The complexity of studies on sequential treatment procedures is evident from more recent investigations in unipolar depression [1, 2]. It is difficult to evaluate complex treatment algorithms in methodologically stringent studies (e.g., randomized, control group studies). The needed high number of cases requires multicentre settings with a large number of sites. But even if enough centers were willing to participate, it would probably be difficult to obtain financial support for the study. Thus, in the long term much of what is actually significantly more important in routine clinical care than the question of whether there is better empirical proof for the efficacy of this or that (licensed!) drug, or whether a drug has better efficacy or tolerability and should therefore be given as the drug of first choice, can possibly not be fully regulated by EBM, or at least only with great difficulty.

After the seemingly positive result of the Texas Medication Algorithm Project [135, 150], a result which can be questioned on the basis of significant baseline differences between the two compared groups, a very comprehensive algorithm study on the treatment of depression took on this challenge: the North American STAR*D study. Despite the impressive sample size, and the excellent planning and organisation of this multi-centre project network, unfortunately only little information on the differential effect of alternative treatments was obtained. Hence, the practical relevance of this study is limited. One of the reasons for this deficiency was that time restraints (continuation of previous treatment instead of switch) were not controlled for [46, 47, 127, 136, 138, 151, 152]. This constellation underlines just how difficult it is to empirically base decisions on complex treatment programs.

Complex clinical decision-making considering medical, health-economic values and individual aspects

The evaluation of comedication or sequential treatment approaches discussed above reveals the complexity of

clinical treatment decisions. But the situation in routine clinical care is more complex since such treatment decisions are normally made on the basis of the special characteristics of each individual patient, his genetic and neurobiological characteristics, previous treatment, psychopathological and other illness characteristics and his disposition to side effects [76, 79, 82, 93].

Health-economic analyses will become increasingly important in face of the shortage of resources in the health system [110, 111, 128]. They can introduce additional aspects into the clinical decision-making process, particularly if the future resources are allocated not primarily on the basis of clinical results of treatment studies but by institutions such as NICE using health-economic parameters to differentiate between different treatment approaches. This approach yields different results depending on the criteria of form and content applied. Physicians should oppose tendencies to prioritize criteria such as hospitalization or inability to work by placing the emphasis on the relevance of the patient's subjective life quality.

Finally, it should be kept in mind that clinical decision-making is not only based on evidence, but also on values. It has been suggested that values-based medicine should be added to EBM [50].

Conclusions

The call for evidence-based treatment in psychiatry is an important movement in the context of the current general demand for EBM. As far as possible, clinical and in particular psychopharmacotherapeutic decision-making processes should be regulated along these lines. This includes the careful processing of the results of randomized clinical studies in narrative, systematic review papers and in statistical meta-analyses. These results can serve as a basis for the preparation of guidelines for therapeutic decision making in routine clinical care, if they are combined with the competence and experience of clinical experts. But a closer look at this principally more meaningful approach reveals that a series of related problems has not yet been adequately solved. Hence, the presumed advantages of this approach carry specific risks which have to be considered as well. It is especially difficult to base decisions or complex therapeutic strategies on evidence data alone. The more evidence is missing, the more it has to be replaced by clinical experience and experts' opinion or concensus. Even Sackett, who is seen as one of the fathers of EBM, understood that EBM should not only rely on results of sophisticated studies, but should also integrate the medical experience of the doctor in charge of an individual patient [140]. Finally, EBM should not exclude a values-oriented medicine.

References

1. Adli M, Rush J, Möller HJ, Bauer M (2003) Algorithms for optimizing the treatment of depression: making the right decision at the right time. Pharmacopsychiatry 36 (Suppl 3):S222–S229
2. Adli M, Bauer M, Rush AJ (2006) Algorithms and collaborative-care systems for depression: are they effective and why? A systematic review. Biol Psychiatry 59:1029–1038
3. Altman DG, Andersen PK (1999) Calculating the number needed to treat for trials where the outcome is time to an event. BMJ 319:1492–1495
4. American Association of Clinical Endocrinologists, American Diabetes Association, American Psychiatric Association, North American Association for the Study of Obesity (2004) Consensus development conference on antipsychotic drugs and obesity and diabetes. Diabetes Care 27:596–601
5. American Psychiatric Association APA Guideline Development Process (2006) http://www.psych.org/psych_pract/treatg/pg/prac_guide.cfm
6. Antes G (2004) The evidence base of clinical practice guidelines, health technology assessments and patient information as a basis for clinical decision-making. Z Ärztl Fortbild Qualitätssich 98:180–184
7. Ascher-Svanum H, Zhu B, Faries D, Ernst FR (2004) A comparison of olanzapine and risperidone on the risk of psychiatric hospitalization in the naturalistic treatment of patients with schizophrenia. Ann Gen Hosp Psychiatry 3:11
8. Ascher-Svanum H, Zhu B, Faries D, Landbloom R, Swartz M, Swanson J (2006) Time to discontinuation of atypical versus typical antipsychotics in the naturalistic treatment of schizophrenia. BMC Psychiatry 6:8
9. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D et al (2004) Grading quality of evidence and strength of recommendations. BMJ 19:1490
10. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B et al (2004) Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 4:38
11. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N et al (2005) Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. BMC Health Serv Res 5:25
12. Baghai TC, Möller HJ, Rupprecht R (2006) Recent progress in pharmacological and non-pharmacological treatment options of major depression. Curr Pharm Des 12:503–515
13. Baldwin DS (2006) Evidence-based guidelines for anxiety disorders: can they improve clinical outcomes? CNS Spectr 11:34–39
14. Baldwin D, Broich K, Fritze J, Kasper S, Westenberg H, Möller HJ (2003) Placebo-controlled studies in depression: necessary, ethical and feasible. Eur Arch Psychiatry Clin Neurosci 253:22–28
15. Bandelow B, Zohar J, Hollander E, Kasper S, Möller HJ, WFSBP Task Force on Treatment Guidelines for Anxiety (2003) World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the pharmacological treatment of anxiety, obsessive–compulsive and posttraumatic stress disorders. World J Biol Psychiatry 3:171–199
16. Bandelow B, Zohar J, Hollander E, Kasper S, Möller HJ, WFSBP Task Force on Treatment Guidelines for Anxiety (2008) World Federation of Societies of Biological Psychiatry

(WFSBP) Guidelines for the pharmacological treatment of anxiety, obsessive–compulsive and post-traumatic stress disorders—first revision. World J Biol Psychiatry 9:248–312
17. Bateman A, Fonagy P (1999) Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. Am J Psychiatry 156:1563–1569
18. Bauer M, Whybrow PC, Angst J, Versiani M, Möller HJ, WFSBP Task Force on Treatment Guidelines for Unipolar Depressive Disorders (2002) World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, Part 1: Acute and continuation treatment of major depressive disorder. World J Biol Psychiatry 3:5–43
19. Bauer M, Whybrow PC, Angst J (2002) World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, Part 2: Maintenance treatment of major depressive disorder and treatment of chronic depressive disorders and subthreshold depressions. World J Biol Psychiatry 3:69–86
20. Bauer M, Bschor T, Pfennig A, Whybrow PC, Angst J, Versiani M, Möller HJ, WFSBP Task Force on Unipolar Depressive Disorders (2007) World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Unipolar Depressive Disorders in Primary Care. World J Biol Psychiatry 8:67–104
21. Bauer M, Tharmanathan P, Volz HP, Möller HJ, Freemantle N (2009) The effect of venlafaxine compared with other antidepressants and placebo in the treatment of major depression: a meta-analysis. Eur Arch Psychiatry Clin Neurosci 259:172–185
22. Beynon S, Soares-Weiser K, Woolacott N, Duffy S, Geddes JR (2009) Pharmacological interventions for the prevention of relapse in bipolar disorder: a systematic review of controlled trials. J Psychopharmacol 23:574–591
23. Bottlender R, Rudolf D, Strauss A, Moller HJ (1998) Antidepressant-associated maniform states in acute treatment of patients with bipolar-I depression. Eur Psychiatry 248:296–300
24. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P (2000) Framework for design and evaluation of complex interventions to improve health. BMJ 321:694–696
25. Carpenter WT (2001) Evidence-based treatment for first-episode schizophrenia? Am J Psychiatry 158:1771–1773
26. Cartabellotta A, Minella C, Bevilacqua L, Caltagirone P (1998) Evidence-based medicine. 3. Systematic reviews: a tool for clinical practice, permanent education and health policy decisions. Italian Group on Evidence-Based Medicine-GIMBE. Recenti Prog Med 89:329–337
27. Cartabellotta A, Montalto G, Notarbartolo A (1998) Evidence-based medicine. How to use biomedical literature to solve clinical problems. Italian Group on Evidence-Based Medicine-GIMBE. Minerva Med 89:105–115
28. Catty J, Burns T, Comas A (2001) Day centres for severe mental illness. Cochrane Database Syst Rev: CD001710
29. Chilvers R, Macdonald GM, Hayes AA (2002) Supported housing for people with severe mental disorders. Cochrane Database Syst Rev: CD000453
30. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JPT, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H et al (2009) Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 373:746–758
31. Citrome L (2007) Show me the evidence: using number needed to treat. South Med J 100:881–884
32. Clark W, Mucklow J (1998) Gathering and weighing the evidence. In: Panton R, Chapman S (eds) Medicines Management. BMJ Books and Pharmaceutical Press, London, pp 59–74
33. Committee for Proprietary Medicinal Products (CPMP) (2002) Note for guidance on clinical investigation of medicinal products in the treatment of depression. http://www.emea.eu.int
34. Correll CU, Rummel-Kluge C, Corves C, Kane JM, Leucht S (2009) Antipsychotic Combinations vs Monotherapy in Schizophrenia: a meta-analysis of randomized controlled trials. Schizophr Bull 35:443–457
35. Craig JC, Irwig LM, Stockler MR (2001) Evidence-based medicine:useful tools for decision making. MJA 174:248–253
36. Crowther R, Marshall M, Bond G, Huxley P (2003) Vocational rehabilitation for people with severe mental illness (Cochrane Review) [computer program]. In: The Cochrane library, issue 3. Update Software, Oxford
37. Cuffel B, Martin J, Joyce AT, Boccuzzi SJ, Loebel A (2006) Lipid and glucose monitoring during atypical antipsychotic treatment: effects of the 2004 ADA/APA consensus statement. Poster presented at the 2006 APA Annual Meeting, Toronto, Canada, 20–25 May 2006
38. Davis JM, Chen N, Glick ID (2003) A meta-analysis of the efficacy of second-generation antipsychotics. Arch Gen Psychiatry 60:553–564
39. Delaney B (2006) Is society losing control of the medical research agenda? BMJ 332:1063–1064
40. Dossenbach M, Erol A, el Mahfoud Kessaci M, Shaheen MO, Sunbol MM, Boland J et al (2004) Effectiveness of antipsychotic treatments for schizophrenia: interim 6-month analysis from a prospective observational study (IC-SOHO) comparing olanzapine, quetiapine, risperidone, and haloperidol. J Clin Psychiatry 65:312–321
41. Eccles M, Mason J (2001) How to develop cost-conscious guidelines. Health Technol Assess 5(16):1–69
42. Eccles M, Grimshaw J, Campbell M, Ramsay C (2003) Research designs for studies evaluating the effectiveness of change and improvement strategies. Qual Saf Health Care 12:47–52
43. Essock SM, Covell NH, Davis SM, Stroup TS, Rosenheck RA, Lieberman JA (2006) Effectiveness of switching antipsychotic medications. Am J Psychiatry 163:2090–2095
44. Falkai P, Wobrock T, Lieberman J, Glenthoj B, Gattaz WF, Moller HJ, WFSBP Task Force on Treatment Guidelines for Schizophrenia (2005) World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia, Part 1: acute treatment of schizophrenia. World J Biol Psychiatry 6:132–191
45. Falkai P, Wobrock T, Lieberman J, Glenthoj B, Gattaz WF, Moller HJ (2006) World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia, Part 2: long-term treatment of schizophrenia. World J Biol Psychiatry 7:5–40
46. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, Quitkin FM, Wisniewski S, Lavori PW, Rosenbaum JF et al (2003) Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. Psychiatr Clin North Am 26:457–494
47. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs M, Luther JF et al (2006) A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. Am J Psychiatry 163:1161–1172
48. Feinstein AR (1995) Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 48:71–79
49. Fritze J, Möller HJ (2001) Design of clinical trials of antidepressants. Should a placebo control arm be included? CNS Drugs 15:755–764
50. Fulford KWM (2004) Facts/values: ten principles of value-based medicine. In: Radden J (ed) The philosophy of psychiatry: a companion. New York, Oxford University Press, pp 205–236

51. Gaebel W, Weinmann S, Sartorius N, Rutz W, McIntyre JS (2005) Schizophrenia practice guidelines: international survey and comparison. Br J Psychiatry 187:248–255
52. Geddes J, Freemantle N, Harrison P, Bebbington P, for the National Schizophrenia Guideline Development Group (2000) Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. BMJ 321:1371–1376
53. Geddes JR, Freemantle N, Mason J, Eccles M, Boynton J (2000) SSRIs versus other antidepressants for depressive disorder. Cochrane Database Syst Rev: CD001851
54. Geddes JR, Calabrese JR, Goodwin GM (2009) Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. Br J Psychiatry 194:4–9
55. Gerson S, Belin TR, Kaufman A, Mintz J, Jarvik L (1999) Pharmacological and psychological treatments for depressed older patients: a meta-analysis and overview of recent findings. Harv Rev Psychiatry 7:1–28
56. Gijsman HJ, Geddes JR, Rendell JM, Nolen WA, Goodwin GM (2004) Antidepressants for bipolar depression: a systematic review of randomized, controlled trials. Am J Psychiatry 161:1537–1547
57. Gonzalez DD (2001) From evidence-based medicine to medicine-based evidence. An Esp Pediatr 55:429–439
58. Gorouhi F, Jafarian S, Firooz A (2007) Reporting of number needed to treat and its difficulties. J Am Acad Dermatol 57:729–730
59. Gray SH (2002) Evidence-based psychotherapeutics. J Am Acad Psychoanal 30:3–16
60. Gray GE, Pinson LA (2003) Evidence-based medicine and psychiatric practice. Psychiatr Q 74:387–399
61. Greenhalgh T (2006) How to read a paper: the basics of evidence based medicine, 3rd edn. BMJ Books/Blackwell, London
62. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L et al (2004) Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 8:iii–iv, 1–72
63. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, Vale L (2006) Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966–1998. J Gen Intern Med 21(Suppl 2):S14–S20
64. Grunze H, Kasper S, Goodwin G, Bowden C, Baldwin D, Licht R, Vieta E, Möller HJ (2002) World Federation of Societies of Biological Psychiary (WFSBP) guidelines for biological treatment of bipolar disorders, Part I: treatment of bipolar depression. World J Biol Psychiatry 3:115–124
65. Grunze H, Kasper S, Goodwin G, Bowden C, Baldwin D, Licht RW, Vieta E, Moller HJ (2003) The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders, Part II: treatment of mania. World J Biol Psychiatry 4:5–13
66. Grunze H, Kasper S, Goodwin G, Bowden C, Moller HJ (2004) The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders, Part III: maintenance treatment. World J Biol Psychiatry 5:120–135
67. Grunze H, Vieta E, Guy M, Goodwin GM, Bowden C, Licht RW, Möller HJ, Kasper S, WFSBP Task Force on Treatment Guidelines for Bipolar Disorders (2009) The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders: update 2009 on the treatment of acute mania. World J Biol Psychiatry. 10:85–116
68. Gupta M (2007) Does evidence-based medicine apply to psychiatry? Theor Med Bioeth 28:103–120
69. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schunemann H (2006) Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians task force. Chest 129:174–181
70. Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schunemann H (2006) An emerging consensus on grading recommendations? ACP J Club 144:A8–A9
71. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336:924–926
72. Hansen R, Gaynes B, Thieda P, Gartlehner G, Deveaugh-Geiss A, Krebs E, Lohr K (2008) Meta-analysis of major depressive disorder relapse and recurrence with second-generation antidepressants. Psychiatr Serv 59:1121–1130
73. Haro JM, Edgell ET, Jones PB, Alonso J, Gavart S, Gregor KJ et al (2003) The European Schizophrenia Outpatient Health Outcomes (SOHO) study: rationale, methods and recruitment. Acta Psychiatr Scand 107:222–232
74. Haro JM, Suarez D, Novick D, Brown J, Usall J, Naber D et al (2007) Three-year antipsychotic effectiveness in the outpatient care of schizophrenia: observational versus randomized studies results. Eur Neuropsychopharmacol 17:235–244
75. Hegerl U, Plattner A, Moller HJ (2004) Should combined pharmaco- and psychotherapy be offered to depressed patients? A qualitative review of randomized clinical trials from the 1990s. Eur Arch Psychiatry Clin Neurosci 254:99–107
76. Henkel V, Seemüller F, Obermeier M, Adli M, Bauer M, Mundt C, Brieger P, Laux G, Bender W, Heuser I et al (2009) Does early improvement triggered by antidepressants predict response/remission? Analysis of data from a naturalistic study on a large sample of inpatients with major depression. J Affect Disord 115:439–449
77. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S (2006) Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. Am J Psychiatry 163:185–194
78. Hirschfeld RMA, Montgomery SA, Aguglia E, Amore M, Delgado PL, Gastpar M, Hawley C, Kasper S, Linden M, Massana J et al (2002) Partial response and nonresponse to antidepressant therapy: current approaches and treatment options. J Clin Psychiatry 63:826–837
79. Holsboer F (2008) How can we realize the promise of personalized antidepressant medicines? Nat Rev Neurosci 8:638–646
80. Jadad AR, Phil D, Cook DJ (1998) Methodology and reports of systematic reviews and meta-analyses-acomparison of Cochrane reviews with articles published in paper-based journals. JAMA 280:278–280
81. Jiang Q, Ahmed S (2009) An analysis of correlations among four outcome scales employed in clinical trials of patients with major depressive disorder. Ann Gen Psychiatry 8:4
82. Juckel G, Pogarell O, Augustin H, Mulert C, Müller-Siechender F, Frodl T, Mavrogiorgou P, Hegerl U (2007) Differential prediction of first clinical response to serotonergic and noradrenergic antidepressants using the loudness dependence of auditory evoked potentials in patients with major depressive disorder. J Clin Psychiatry 68:1206–1212
83. Kallert TW (2005) [Is mental health services research in need of randomised controlled trials?] Braucht psychiatrische Versorgungsforschung randomisierte kontrollierte Studien? Psychiatr Prax 32:375–377

84. Kane JM, Honigfeld G, Singer J, Meltzer H (1988) Clozapine in treatment-resistant schizophrenics. Psychopharmacol Bull 24:62–67
85. Kane JM, Honigfeld G, Singer J, Meltzer HY (1988) Clozapine for the treatment-resistant schizophrenic: a double-blind comparison versus chlorpromazine/benztropine. Arch Gen Psychiatry 48:789–796
86. Kawamura T, Tamakoshi A, Wakai K, Ohno Y (1999) Evidence-based medicine and 'The Cochrane Collaboration'. Nippon Koshu Eisei Zasshi 46:498–506
87. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. PLoS Med 5:260–268
88. Kjaergard LL, Is-Nielsen B (2002) Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ 325:249
89. Klein DF (2000) Flawed meta-analyses comparing psychotherapy with pharmacotherapy. Am J Psychiatry 157:1204–1211
90. Kunz R, Vist G, Oxman A (2007) Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 18(2):MR000012
91. Lehman AF, Steinwachs DM (1998) Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. Schizophr Bull 24:1–10
92. Leichsenring F, Rabung S, Leibing E (2004) The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: a meta-analysis. Arch Gen Psychiatry 61:1208–1216
93. Lerer B, Macciardi F (2002) Pharmacogenetics of antidepressant and mood-stabilizing drugs: a review of candidate-gene studies and future research directions. Int J Neuropsychopharmacol 5:255–275
94. Leucht S (2004) Amisulpride a selective dopamine antagonist and atypical antipsychotic: results of a meta-analysis of randomized controlled trials. Int J Neuropsychopharmacol 7:S15–S20
95. Leucht S, Wahlbeck K, Hamann J, Kissling W (2003) New generation antipsychotics versus low-potency conventional antipsychotics: a systematic review and meta-analysis. Lancet 361:1581–1589
96. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM (2009) How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Mol Psychiatry 14:429–447
97. Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM (2009) Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet 373:31–41
98. Leucht S, Komossa K, Rummel-Kluge C, Corves C, Hunger H, Schmid F, Lobos CA, Schwarz S, Davis JM (2009) A meta-analysis of head-to-head comparisons of second-generation antipsychotics in the treatment of schizophrenia. Am J Psychiatry 166:152–163
99. Lieberman JA, Greenhouse J, Hamer RM, Krishnan KR, Nemeroff CB, Sheehan DV, Thase ME, Keller MB (2005) Comparing the effects of antidepressants: consensus guidelines for evaluating quantitative reviews of antidepressant efficacy. Neuropsychopharmacology 30:445–460
100. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, Keefe RS, Davis SM, Davis CE, Lebowitz BD et al (2005) Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med 353:1209–1223
101. Linden M (2005) Leitlinien und die Psychologie medizinischer Entscheidungsprozesse bei der Behandlung depressiver Erkrankungen [Guidelines and the psychology of medical decision making in the treatment of depressive disorders]. Fortschr Neurol Psychiat 73:249–258
102. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, Korslund KE, Tutek DA, Reynolds SK, Lindenboim N (2006) Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 63:757–766
103. Lohr KN (2004) Rating the strength of scientific evidence: relevance for quality improvement programs. Int J Qual Health Care 16:9–18
104. Maier W, Möller HJ (2005) Metaanalyses—highest level of empirical evidence? Eur Arch Psychiatry Clin Neurosci 255:369–370
105. Maier W, Möller HJ (2008) Meta-analyses: a method to maximize the evidence from clinical studies? Eur Arch Psychiatry Clin Neurosci (in preparation)
106. Manser R, Walters EH (2001) What is evidence-based medicine and the role of the systematic review: the revolution coming your way. Monaldi Arch Chest Dis 56:33–38
107. Marshall M, Lockwood A (2003) Assertive communication treatment for people with severe mental disorder (Cochrane Review) [computer program]. In: The Cochrane library, issue 3. Update Software, Oxford
108. Marshall M, Gray A, Lockwood A, Green R (2003) Case management for people with severe mental disorders (Cochrane Review) [computer program]. Update Software, Oxford
109. Marshall M, Crowther R, Almaraz-Serrano A, Creed F, Sledge W, Kluiter H, Roberts C, Hill E, Wiersma D (2003) Day hospital versus admission for acute psychiatric disorders. Cochrane Database Syst Rev: CD004026
110. McCabe C, Dixon S (2000) Testing the validity of cost-effectiveness models. Pharmacoeconomics 17:501–513
111. McCabe C, Claxton K, Culyer AJ (2008) The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics 26:733–744
112. Melander H, Ahlqvist-rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326:1171–1173
113. Melander H, Salmonson T, Abadie E, van Zwieten-Boot B (2008) A regulatory Apologia—a review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. Eur Neuropsychopharmacol 18:623–627
114. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 354:1896–1900
115. Möller HJ (2000) State of the art of drug treatment of schizophrenia and the future position of the novel/atypical antipsychotics. World J Biol Psychiatry 1:204–214
116. Möller HJ (2006) Ethical aspects of publishing. World J Biol Psychiatry 7:66–69
117. Möller HJ (2008) Do effectiveness ("real world") studies on antipsychotics tell us the real truth? Eur Arch Psychiatry Clin Neurosci 258:257–270
118. Möller HJ (2008) Isn't the efficacy of antidepressants clinically relevant? A critical comment on the results of the metaanalysis by Kirsch et al. 2008. Eur Arch Psychiatry Clin Neurosci 258:451–455
119. Möller HJ (2009) Declaration of conflicts of interest in scientific publications. World J Biol Psychiatry 10:2–5 Editorial
120. Möller HJ (2009) Standardised rating scales in Psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. World J Biol Psychiatry 10:6–26
121. Möller HJ, Grunze H (2000) Have some guidelines for the treatment of acute bipolar depression gone too far in the

restriction of antidepressants? Eur Arch Psychiatry Clin Neurosci 250:57–68
122. Möller HJ, Grunze H, Broich K (2006) Do recent efficacy data on the drug treatment of acute bipolar depression support the position that drugs other than antidepressants are the treatment of choice? A conceptual review. Eur Arch Psychiatry Clin Neurosci 256:1–16
123. Möller HJ, Langer S, Schmauss M (2007) Escitalopram in clinical practice: results of an open-label trial in outpatients with depression in a naturalistic setting in Germany. Pharmacopsychiatry 40:53–57
124. Möller HJ, Baldwin D, Goodwin G, Kasper S, Okasha A, Stein DJ, Tandon R, Versiani M (2008) Do SSRIs or antidepressants in general increase suicidality? WPA Section on Pharmacopsychiatry: consensus statement. Eur Arch Psychiatry Clin Neurosci 258:3–23
125. Nemeroff CB, Entsuah R, Benattia I, Demitrack M, Sloan DM, Thase ME (2008) Comprehensive analysis of remission (COMPARE) with venlafaxine versus SSRIs. Biol Psychiatry 63:424–434
126. NICE National Institute for Health and Clinical Excellence Anxiety (amended): Management of Anxiety (Panic Disorder, with or without Agoraphobia, and Generalised Anxiety Disorder) in Adults in Primary, Secondary and Community Care. (2007) CG22
127. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G et al (2006) A comparison of lithium and T(3) augmentation following two failed medication treatments for depression: a STAR*D report. Am J Psychiatry 163:1519–1530
128. O'Hagan A, McCabe C, Akehurst R, Brennan A, Briggs A, Claxton K, Fenwick E, Fryback D, Sculpher M, Spiegelhalter D et al (2005) Incorporation of uncertainty in health economic modelling studies. Pharmacoeconomics 23:529–536
129. Ollenschlager G, Berenbeck C, Low A, Stobrawa F, Kolkmann FW (2002) The National Program for disease management—guidelines under the auspices of the German Medical Association—a method report. Z Arztl Fortbild Qualitatssich 96:545–548
130. Patsopoulos NA, Ioannidis JP, Analatos AA (2006) Origin and funding of the most frequently cited papers in medicine: database analysis. BMJ 332:1061–1064
131. Pilling S, Price K (2006) Developing and implementing clinical guidelines: lessons from the NICE schizophrenia guideline. Epidemiol Psichiatr Soc 15:109–116
132. REJr Kelly, Cohen LJ, Semple RJ, Bialer P, Lau A, Bodenheimer A et al (2006) Relationship between drug company funding and outcomes of clinical psychiatric research. Psychol Med 36:1647–1656
133. Riedel M, Strassnig N, Müller N, Zwack P, Möller HJ (2005) How representative of everyday clinical populations are schizophrenia patients enrolled in clinical trials? Eur Arch Psychiatry Clin Neurosci 255:143–148
134. Rush AJ (2007) STAR*D: What have we learned? Am J Psychiatry 164:201–204
135. Rush JA, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Trivedi MH, Suppes T, Miller AL, Biggs MM, Shores-Wilson K et al (2003) Texas Medication Algorithm Project, Phase 3 (TMAP-3): rationale and study design. J Clin Psychiatry 64:357–369
136. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM et al (2004) Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. Control Clin Trials 25:119–142
137. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD et al (2006) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 163:1905–1917
138. Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, Ritz L, Biggs MM, Warden D, Luther JF et al (2006) Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. N Engl J Med 354:1231–1242
139. Sackett DL (2000) Evidence-based medicine: how to practice and teach EBM. Churchill Livingstone, New York
140. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312:71–72
141. Sheehan DV, Burnham DB, Iyengar MK, Perera P, for the Paxil CR Panic Disorder Study Group (2005) Efficacy and tolerability of controlled-release paroxetine in the treatment of panic disorder. J Clin Psychiatry 66:34–40
142. Sheehan DV, Keene MS, Eaddy M, Krulewicz S, Kraus JE, Carpenter DJ (2008) Differences in medication adherence and healthcare resource utilization patterns. older versus newer antidepressant agents in patients with depression and/or anxiety disorders. CNS Drugs 22:963–973
143. Stauffer V, Karagianis J, Sutton V, Ascher-Svanum H, Treuer T, Silva de Lima M, Ball T, Poole-Hoffmann V, Tohen M (2008) Number Needed to Treat (NNT) and Number Needed to Harm (NNH) in randomized, blinded trials comparing olanzapine to other atypical antipsychotics for treatment of schizophrenia. Clin Schizophr Relat Psychoses: 136–146
144. Sterne JA, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 53:1119–1129
145. Stone JM, Davis JM, Leucht S, Pilowsky LS (2009) Cortical dopamine D2/D3 receptors are a common site of action for antipsychotic drugs—an original patient data meta-analysis of the SPECT and PET in vivo receptor imaging literature. Schizophr Bull 35:789–797
146. Stroup TC (2007) Heterogeneity of treatment effects in schizophrenia. Am J Med 120:526–531
147. Stroup TS, Lieberman JA, McEvoy JP, Davis SM, Swartz MS, Keefe RS, Miller AL, Rosenheck RA, Hsiao JK, For the CATIE Investigators (2009) Results of phase 3 of the CATIE schizophrenia trial. Schizophr Res 107: 1–12
148. Swartz MS, Stroup TS, McEvoy JP, Davis SM, Rosenheck RA, Keefe RS, Hsiao JK, Lieberman JA (2008) What CATIE found: results from the schizophrenia trial. Psychiatr Serv 59:500–506
149. Tandon R, Belmaker RH, Gattaz WF, Lopez-Ibor JJ Jr, Okasha A, Singh B, Stein DJ, Olie JP, Fleischhacker WW, Möller HJ et al (2008) World Psychiatric Association Pharmacopsychiatry Section statement on comparative effectiveness of antipsychotics in the treatment of schizophrenia. Schizophr Res 100:20–38
150. Trivedi MH, Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Key T, Biggs MM, Shores-Wilson K, Witte B et al (2004) Clinical results for patients with major depressive disorder in the Texas Medication Algorithm Project. Arch Gen Psychiatry 61:669–680
151. Trivedi MH, Fava M, Wisniewski SR, Thase ME, Quitkin F, Warden D, Ritz L, Nierenberg AA, Lebowitz BD, Biggs MM et al (2006) Medication augmentation after the failure of SSRIs for depression. N Engl J Med 354:1243–1252
152. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ et al (2006) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 163:28–40

153. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publications of antidepressant trials and its influence on apparent efficacy. N Eng J Med 358:252–260
154. Wampold BE, Minami T, Baskin TW, Callen TS (2002) A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. J Affect Disord 68:159–165
155. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR (2007) The STAR*D project results: a comprehensive review of findings. Curr Psychiatry Rep 9:449–459
156. Wijkstra J, Lijmer J, Balk FJ, Geddes JR, Nolen WA (2006) Pharmacological treatment for unipolar psychotic depression: systematic review and meta-analysis. Br J Psychiatry 188:410–415
157. Yank V, Rennie D, Bero LA (2007) Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. BMJ 335:1202–1205