Effectiveness studies: advantages and disadvantages

Hans-Jürgen Möller, MD, PhD

In recent years, so-called "effectiveness studies," also called "real-world studies" or "pragmatic trials," have gained increasing importance in the context of evidence-based medicine. These studies follow less restrictive methodological standards than phase III studies in terms of patient selection, comedication, and other design issues, and their results should therefore be better generalizable than those of phase III trials. Effectiveness studies, like other types of phase IV studies, can therefore contribute to knowledge about medications and supply relevant information in addition to that gained from phase III trials. However, the less restrictive design and inherent methodological problems of phase IV studies have to be carefully considered. For example, the greater variance caused by the different kinds of confounders as well as problematic design issues, such as insensitive primary outcome criteria, unblinded treatment conditions, inclusion of chronic refractory patients, etc, can lead to wrong conclusions. Due to these methodological problems, effectiveness studies are on a principally lower level of evidence, adding only a complementary view to the results of phase III trials without falsifying their results.

Dialogues Clin Neurosci. 2011;13:199-207.

In the context of evidence-based medicine,[1] randomized control-group trials (RCTs) are considered to be the decisive level of scientifically proven evidence as far as therapeutic aspects are concerned.[2] Placebo-controlled trials, especially for certain psychiatric indications, are ranked higher in terms of evidence than active control-group studies.[3] Especially in terms of licensing perspectives, there is a demand from the European Medicines Agency and the Food and Drug Administration to demonstrate efficacy based on RCTs including a placebo control group for obvious methodological reasons. The knowledge gained from noninterventional (observational) studies (NIS) as well as from single-case studies is only seen as being relevant when it is an addition to such studies or a replacement in indications 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.[4,5] Thus, randomized control-group studies have a higher value than nonrandomized or uncontrolled studies.

Keywords: *effectiveness study; real-world study; antipsychotic; first-generation antipsychotic (FGA); second-generation antipsychotic (SGA)*

Author affiliations: Department of Psychiatry, Ludwig-Maximilians-University München, Munich, Germany

Address for correspondence: Prof Dr Hans-Jürgen Möller, Department of Psychiatry, Ludwig-Maximilians-University München, Nussbaumstrasse 7, 80336 Munich, Germany
(e-mail: hans-juergen.moeller@med.uni-muenchen.de)

Do effectiveness studies tell us the truth?

There is a general consensus that the results of phase III studies are not fully generalizable: they have a high internal validity but insufficient external validity. One of the reasons for this is the strict selection of patients according to various clinically relevant characteristics such as the exclusion of suicidality, comorbidity, etc. For this reason it has long been a tradition within clinical psychopharmacology to complement the phase III trial results with ones more strongly oriented towards everyday clinical practice and conditions, ie, studies in patients who better represent the "average" patients and treated under conditions as close as possible to "routine" care, eg, phase IV studies *(Figure 1)*. However, it has thereby always been stressed that because of many immanent methodological problems, eg, biases due to lack of double-blind conditions or any blinding, such as phase naturalistic observational studies (NIS), only deliver complementary knowledge and cannot falsify the results of phase III studies.[6]

However, this strict rule can be weakened if the phase IV studies are performed, like phase III studies, as randomized control-group studies in an unblinded or even in blind or double-blind approach.

Some experts seem inclined to attach a greater importance to the results of these studies than to the methodologically stricter phase III studies.[7] This might in particular be the result from criticism arising from the increasingly common practice, especially in the USA, to include, in phase III studies, not "real" patients from care settings, but suitable persons found through advertisements. Of course, rather than this questionable approach, properly performed phase III studies in "real" patients should be advocated. Even so, some experts judge the "real-world approach" of effectiveness studies to be more valuable than phase III trials, at least in terms of clinical relevance.

Some methodological considerations on effectiveness studies

Effectiveness studies are intended to fill the gap between methodologically rigorous RCTs in the sense of phase III trials and naturalistic observational studies. As such, they are hybrids of the RCT methodology and naturalistic designs and are therefore termed "practical clinical trials." [8] They are intentionally designed to evaluate the effectiveness of the treatments under real-world conditions and in patient samples representative of everyday clinical practice *(Table 1)*. They can be performed as RCTs, but less demanding designs are also possible. If they use even a blind[9] or double-blind[10] RCT approach they come close to phase III trials considering design aspects, with the only difference being that patient selection is not that restrictive and that, eg, comorbidity or comedication are allowed.

In order to avoid guidelines completely losing their rela-

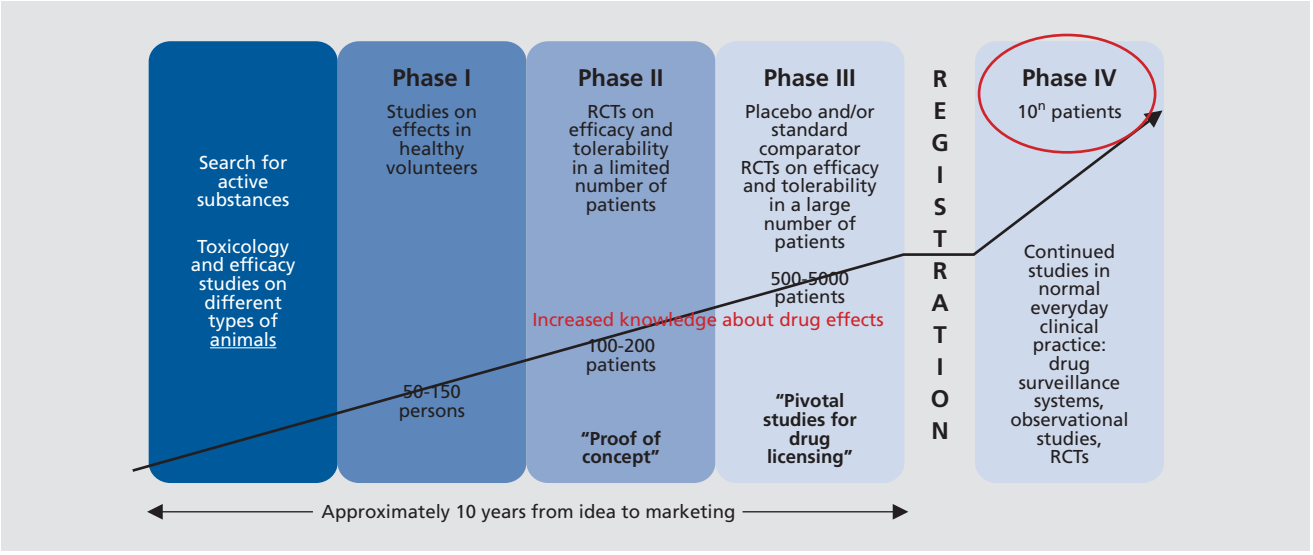

Figure 1. The 4-phase model of clinical psychopharmacology. RCT, randomized controlled trial

Clinical trials of "effectiveness"
More relaxed exclusion criteria, permitting wider range of:
Patients (eg, comorbidity not excluded)
Treatment settings and interventions (including adjunctive treatments)
Emphasis on clinical need to determine treatment doses, etc
Levels and/or types of psychopathology
Forms of outcome criteria, such as:
Time to discontinuation
Quality of life
Preference of self-rating instruments or global ratings
Advantages
Higher external validity
Arguably greater applicability to "real-world" practice settings
Capacity to inform policy process
Longer duration can be easier achieved
Can enrol large number of patients more easily
Disadvantages
Internal validity limited
Cannot be used to examine effective dose ranges
Cannot make as meaningful clinical comparisons between agents
Clinical trials of "efficacy"
Highly restricted inclusion criteria to reduce confounding biases
Randomization and blinding, also to reduce bias
Treatment driven exclusively by study protocol
Patients remain only in the treatment group originally assigned
Fewer treatment adjustments are allowed
Strict limitations on adjunctive treatment
Measures taken to insure all members of treatment group receive same intervention(s)
Use of well-validated outcome assessment
Advantages
Higher internal validity for clinical effects
Higher internal validity for adverse effects, tolerability
Contextual and human factors controlled for
Considered "best quality" clinical evidence for informing treatment decisions
Disadvantages
Stringent inclusion criteria limit external validity
Outcome measures may not reflect crucial advantages and limitations of interventions being studied
Outcome measures may not address issues most important to patients and families
Often short in duration

Table I. Some characteristics of clinical trials of "efficacy" vs trials of "effectiveness."

tionship with clinical reality by preferring study types with too little generalizability, 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, eg, randomized control-group studies versus a standard drug; the results should at least show a tendency towards consistency. The 3-arm study design recommended by the European regulatory authority, EMEA/CPMP,[11] in which the experimental substance is compared with 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. Therefore, other types of studies traditionally considered to be phase IV should be part of the evaluation process.

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 additionally be obtained from phase IV studies, which are more closely oriented towards routine clinical care,[12-17] 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 (eg, placebo-controlled studies) without ascertaining whether consistent results are available from less restrictive but more generalizable study types. A future grading of evidence that is more relevant for clinical reality should assess whether results are available from studies with both high internal (eg, control-group studies) and high external (eg, effectiveness studies, observational studies) validity and whether the results are principally congruent. So far, the current interest in effectiveness studies is principally positive.[10,18,19] However, the results of these effectiveness studies should not be overinterpreted due to their principal methodological limitations (as demonstrated, eg, for the Clinical Antipsychotic Trials of Intervention Effectiveness [CATIE] trial).[6]

The inclusion of "confounders" (from the perspective of a phase III trial) such as comorbidity or comedication increases the variance and results in a reduced signal-to-noise ratio, which makes it more difficult to find differences between two groups (β error problem), even if these factors are adequately considered in the statistical analysis. It might sometimes even be difficult to judge without placebo conditions whether there is a real drug effect, especially if the pre-post difference is unexpectedly low and if there are no differences between two active comparators. Given the fact that these pragmatic trials mostly compare two active compounds, it should be accepted on the basis of the traditional methodology of clinical psychopharmacological trials that only proof of superiority in the statistical sense counts, while the failure to demonstrate a statistically significant difference cannot be interpreted as showing that both treatments are comparable.[3] The latter conclusion is not permissible for principal methodological reasons.

A different statistical design is required to demonstrate equivalency: the so-called equivalency design. However, this methodological approach is also far from the unambiguity of superiority trials. For example, without a placebo control, which is characteristic for effectiveness studies,[20-23] one cannot be sure that the active drugs are being compared in a drug-sensitive sample *(Table II)*.[3] The worst-case scenario is that the drugs show no outcome difference because they are not effective at all in the respective sample. This is not as unlikely as some might believe. In the field of antidepressants, failed studies—in the sense that in a 3-arm study comparing an experimental drug with a standard comparator and placebo not even the standard comparator (internal validator) differs from placebo—are quite common.[24] In recent years there has even been an increasing number of failed studies, especially in the United States, not only in the field of antidepressants but also in the field of antipsychotics, although the antipsychotics generally have a larger effect size than antidepressants. Several factors are relevant in this context, such as low inter-rater reliability, especially in huge multicenter trials, inclusion of less responsive patients, more chronic patients with residual symptomatology or comorbid patients, no restriction of permitted comedications, etc.

In discussing methodological aspects of effectiveness studies it should be questioned whether outcome criteria such as "nondiscontinuation," or similar categorical end points like "level of caring," preferably applied in some effectiveness studies, really are ideal outcome criteria, given the fact that they can easily be influenced by the investigators (who may be biased by their expectations if they are not blinded) and are of poorer psychometric value than dimensional ones.

It can be generally questioned whether "nondiscontinuation" really reflects only efficacy and tolerability aspects, or whether other parameters beyond drug effects are also involved, eg, confidence in the therapeutic concept. For example, therapeutic concepts like psychotherapy, herbal drug therapy, etc, might be more acceptable to a subgroup of patients, although they may have a lower level of efficacy. Different aspects of tolerability can have different effects on discontinuation, depending on the specific tolerability problems and on the time patterns of side effects. Thus, one can presume

	Advantages	Disadvantages
Placebo-controlled studies	Allow estimation of the assay sensitivity and thus internal validation of the study	Perhaps higher risk from "nontreatment"
	Allow better evaluation of the clinical relevance	Perhaps more limited generalisability of the results to the general population
	Smaller sample size	
	Lower study costs	
Studies with an active control	Supply data on relative efficacy and tolerability	Risk of false studies because assay sensitivity is lacking
	At least theoretically no inactive treatment	Equivalence/noninferiority not suitable as proof of efficacy
	Fewer dropouts due to lack of efficacy	Active comparator may not be standard therapy
	May be more acceptable to an ethics commission	More dropouts due to adverse events
		Tendancy to minimize efficacy differences
		Larger sample sizes
		Higher study costs

Table II. Advantages and disadvantages of using an active control or placebo in clinical studies.

that severe extrapyramidal symptoms occurring right at the start of a study result in an early dropout, the slow development of weight gain rather a later dropout, and tardive dyskinesia (TD) or in most cases even metabolic disorder, a much later dropout. This means that a rough measurement like "discontinuation" or "time to discontinuation" causes a biased distortion per se with respect to the individual antipsychotics being evaluated. This becomes even worse if the transition from the pretreatment antipsychotic to the study antipsychotic is taken into consideration, in particular if it is direct, without a sufficiently long washout phase. Depending on the pharmacological profile of the respective pretreatment drug, for example in terms of D_2 potency, anticholinergic or antihistaminergic properties, and the related pharmacological profile of the study drug, several problems can appear immediately after transition.[25] These can include reduced antipsychotic efficacy, discontinuation symptoms, hangover of side effects wrongly attributed to the study drug, pharmacodynamic interactions in terms of oversedation, histaminergic, or cholinergic rebound phenomena, etc. Thus, there are good and bad combinations of drugs for this transition process. Theoretically, the best transition is one in which the pretreatment and the study drug are identical. There are also other critical issues that need to be considered in this context.[26,27]

Quality of life

Another preferred measure of global outcome used as a primary outcome criterion in some effectiveness studies is "quality of life." There is no doubt that this is an important outcome criterion which reflects the subjective dimension of the patient's experience.[28-30] The classical approach in quality of life research assesses quality of life using a self-rating scale in order to guarantee the subjective perspective. The SF36[31,32] is particularly widely used in psychiatry as well as in other fields of medicine, but there are also several other scales to assess this dimension.[33-35] This leads to the general problem of self-rating approaches for the assessment of the primary outcome, if they are not complemented by an observer rating approach. For example, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study[18] widely relies on self-rating results to assess outcome in terms of depression severity.[9]

Generally, there are pros and cons for the use of self-rating scales. They give a complementary view to the observer rating of the same construct/dimension.[36,37] The correlation between the observer ratings and self-ratings might not be high and may be quite changeable, depending on the psychopathological state in terms of severity and type of symptoms.[38] It is often unclear exactly what self-ratings of quality of life reflect; severity of the psychopathological state in the global sense, certain dimensions of the psychopathological state, eg, depression, current mood more than real depressive symptoms, side effects of drugs, or the psychosocial situation.[29,39-43] If such a scale is used as the primary outcome criterion of a study, it is doubtful whether it is sensitive enough to detect intergroup differences in treatment-induced changes, given the high variance of self-rating in general and of self-ratings of quality of life in particular. For example, not many of the studies on antipsychotics that used a quality of life scale as a secondary outcome criterion found significant intergroup differences.[29,29] Thus, the use of a quality of life scale carries a high risk of not finding significant differences between two drugs, especially if both are active drugs.

Do effectiveness studies generally fulfil their claim of treating less selective samples of patients than phase III studies? At least some apparently do not. For example, in the effectiveness study comparing olanzapine and haloperidol in the treatment of schizophrenia,[44] of the 4386 patients assessed for eligibility, only 309 were included in the study (7.0%). This rate is even somewhat lower than the usual rate of 10% to 15% in phase III studies.[45] Some effectiveness studies appear to have a different kind of selection of patients than phase III trials. Often, patients with milder and more chronic symptoms may be selected than is the case in phase III studies, thus making it more difficult per se to demonstrate drug effects and in particular differences between drug effects, because a relevant subgroup of patients might be partially unresponsive to a drug. The data from the Cost Utility of the Latest Antipsychotics in Severe Schizophrenia (CUtLASS) study serve as an example here. In this study, the pre-post changes in the Positive And Negative Symptom Scale (PANSS) positive score after 52 weeks amounted to only 2.0 in the first-generation antipsychotic (FGA) arm and 1.5 in the second-generation antipsychotic (SGA) arm; these changes are extremely low, even when one takes into account that this study was not an acute treatment study but rather a switch study in partially improved/stabilized patients. Also CATIE[46] and STAR*D[47] patients seem to be more on the chronic and even partially refractory pole.

In order to understand some of the methodological problems of "effectiveness" studies in more detail, the respective review by Möller on effectiveness studies in the field of antipsychotics[6] should be taken into consideration. It is interesting that some of these studies were published in high-ranking journals, although some of them have considerable methodological shortcomings which mean that the conclusions drawn are not tenable, especially not when they are used to falsify the results of phase III studies. Most of these studies arrived at the result that SGAs were generally not superior to FGAs and are thus faced with the comment that not proving superiority does not mean equivalence. The EUFEST study was the only able to demonstrate superiority of SGAs vs haloperidol. A finding of superiority is, for principal methodological reasons (see above) more valid, especially when considering the increased number of confounders in effectiveness studies, than the finding of no statistical differences, which is always difficult to interpret.

The CATIE study

The most famous of effectiveness studies on antipsychotics is the CATIE study.[10] There is no doubt that the CATIE study is an important study when one considers, for example, the large sample size (N=1493 in 57 centers), the complex design with several parallel treatment arms, the 18-month duration of treatment of the first phase, inclusion of sequential treatment phases, etc (phase 1 of the study was published in 2005[10]). Also, the double-blind conditions of this study and the sophisticated and comprehensive statistical analysis of the extensive database are appealing. The study has received a lot of publicity, particularly in the general press, where it was portrayed as showing that SGAs are for the most part not better, but much more expensive, than FGAs. This conclusion is not tenable because of the methodological failings described above and elsewhere.[6,48,49] However, to end on a more positive note, many other results not only from phase 1 but also phase 2 and 3 are of relevance for clinicians, eg, on different side-effect patterns of individual SGAs, on metabolic issues, on meaningful sequences of antipsychotic treatment in case of partial nonresponse, on the unique efficacy of clozapine in refractory patients, etc.[46,50]

In the field of antidepressants there are not so many effectiveness studies. To mention one there is the "Texas Algorithm Study" which tried to demonstrate the superiority of the algorithm approach in treating depressive patients by comparing treatment outcome of depressive patients from two different hospitals. The outcome was more advantageous in the hospital where the algorithm had been applied. However, the weakness of this study was the baseline differences in the two samples, indicating that the patients in the algorithm sample probably had a more positive prognosis. Two other studies which evaluated the algorithm approach in a "real-world" RCT could confirm the superiority of the treatment strategy.[51,52]

The most famous effectiveness study in the field of depression treatment is the STAR*D study.[53] Even more than the CATIE study, this study was a gigantic endeavor in terms of sample size, complexity in design, etc. It investigated under unblinded conditions two different sequential treatment approaches in depressive outpatients, who were randomized at baseline to two different groups. At each level of the complex treatment algorithm the outcome difference between the different groups were evaluated. The methodological problems of this study include the low Hamilton Depression Rating Scale (HAMD) inclusion criteria (HAMD >14), the recruitment of more or less chronic patients in poor psychosocial conditions, overly optimistic power calculations with the consequence that latest for level 3 and 4 the study did not have the necessary power to detect clinically relevant differences. None of the different drug treatment approaches on each level of the sequential treatment algorithm was statistically superior to any of the others; at most some showed a numerical degree of superiority. This "real-world" study reached no clear efficacy results due to inherent methodological problems. From a statistical point of view it does not seem unproblematic that eg, the STAR*D study data were used to generate about 100 publications answering different questions, each of which reporting results based on multiple testings. Given all these problems it has to be questioned whether many really clinically relevant conclusions can be drawn from this study.

Of special methodological interest is the finding that the outcome difference between an a posteriori defined efficacy sample and an effectiveness sample was not as huge as hypothesized.[54] This finding was supported by the results of a naturalistic study on about 1000 depressive inpatients where a similar approach of subdividing the sample a posteriori had been applied.[55] These findings

underline that although there are differences in the sample characteristics of phase III trials and "real-world" trials,[56] the relevance for a different outcome does not have to be as huge as anticipated. Thus, phase III studies are apparently more than only "proof of concept" studies, but have some, although limited, generalizability for real-world patients.

Summary and conclusions

Effectiveness studies can contribute to our knowledge about the use and effectiveness of medications. They help to understand that even novel/expensive drugs have their limitations and that it may not be possible to demonstrate consistently their hypothesized superiority in terms of efficacy, safety, compliance, quality of life, etc under "real-world" conditions in chronic, partially refractory, or comorbid patients. In general they can also supply interesting data on dosing issues, sequences of drugs in case of partial response and side-effect patterns.

Altogether, the effectiveness studies seem to have a lot of methodological problems, making it difficult to interpret their results. Given the fact that increased variance due to the inclusion of chronic/poorly responsive/comorbid patients, insensitive or problematic outcome parameters, and inadequate sample size increase the risk of a β-error (failure to detect a difference although there is one), and that unblinded designs can induce different kinds of biases. Caution has to be applied when interpreting the results of trials with such problems.

In addition, it is questionable whether some effectiveness studies really do represent the real-world treatment situation better than classical acute and long-term phase III studies, as some of them obviously also recruit a selective patient sample, although the selection is of a different kind than in phase III studies. Effectiveness studies can therefore give only a complementary and not a superior picture of reality. Effectiveness studies, especially those with an inadequate experimental design, are definitely not suitable to cast doubt on the results of the methodologically much stricter phase III studies. ❑

REFERENCES

1. Möller HJ, Maier W. Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations. *Eur Arch Psychiatry Clin Neurosci*. 2010;260:25-39.
2. Kunz R, Vist G, Oxman A. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. 2007:MR000012.
3. Möller HJ, Broich K. Principle standards and problems regarding proof of efficacy in clinical psychopharmacology. *Eur Arch Psychiatry Clin Neurosci*. 2010;260:3-16.
4. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321:694-696.
5. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003;12:47-52.
6. Möller HJ. Do effectiveness ("real world") studies on antipsychotics tell us the real truth? *Eur Arch Psychiatry Clin Neurosci*. 2008;258:257-270.
7. Lieberman JA, Greenhouse J, Hamer RM, et al. Comparing the effects of antidepressants: consensus guidelines for evaluating quantitative reviews of antidepressant efficacy. *Neuropsychopharmacology*. 2005;30:445-460.
8. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290:1624-1632.
9. Khan A. Suicide rates of clinical trials of SSRIs, other antidepressants and placebo: analysis of FDA reports. *Am J Psychiatry*. 2003;160:790-792.
10. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209-1223.
11. Committee for Proprietary Medicinal Products (CPMP). Note for guidance on clinical investigation of medicinal products in the treatment of depression. Available at: http://www.emea.eu.int.
12. Ascher-Svanum H, Zhu B, Faries D, Ernst FR. A comparison of olanzapine and risperidone on the risk of psychiatric hospitalization in the naturalistic treatment of patients with schizophrenia. *Ann Gen Hosp Psychiatry*. 2004;3:11.
13. Ascher-Svanum H, Zhu B, Faries D, Landbloom R, Swartz M, Swanson J. Time to discontinuation of atypical versus typical antipsychotics in the naturalistic treatment of schizophrenia. *BMC Psychiatry*. 2006;6:8.
14. Dossenbach M, Erol A, el Mahfoud Kessaci M, et al. 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*. 2004;65:312-321.
15. Haro JM, Edgell ET, Jones PB, et al. The European Schizophrenia Outpatient Health Outcomes (SOHO) Study: rationale, methods and recruitment. *Acta Psychiatr Scand*. 2003;107:222-232.
16. Haro JM, Suarez D, Novick D, et al. Three-year antipsychotic effectiveness in the outpatient care of schizophrenia: observational versus randomized studies results. *Eur Neuropsychopharmacol*. 2007;17:235-244.
17. Möller HJ, Langer S, Schmauss M. Escitalopram in clinical practice: results of an open-label trial in outpatients with depression in a naturalistic setting in Germany. *Pharmacopsychiatry*. 2007;40:53-57.
18. Rush AJ, Fava M, Wisniewski SR, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119-142.
19. Sheehan DV, Keene MS, Eaddy M, Krulewicz S, Kraus JE, Carpenter DJ. Differences in medication adherence and healthcare resource utilization patterns. older versus newer antidepressant agents in patients with depression and/or anxiety disorders. *CNS Drugs*. 2008;22:963-973.
20. Adam D, Kasper S, Möller HJ, Singer EA. Placebo-controlled trials in major depression are necessary and ethically justifiable: how to improve the communication between researchers and ethical committees. *Eur Arch Psychiatry Clin Neurosci*. 2005;255:258-260.
21. Baldwin D, Broich K, Fritze J, Kasper S, Westenberg H, Möller HJ. Placebo-controlled studies in depression: necessary, ethical and feasible. *Eur Arch Psychiatry Clin Neurosci*. 2003;253:22-28.
22. Möller HJ. Sind placebokontrollierte Studien zum Wirksamkeitsbeweis von Antidepressiva notwendig? *Nervenarzt*. 2004;5:421-424.
23. Möller HJ. Plazebo-kontrollierte Studien zum Wirkungsnachweis von Antidepressiva sind notwendig! *Psychopharmakotherapie*. 2003;10:85-86.

Estudios de eficacia: ventajas y desventajas

En los últimos años los así llamados "estudios de eficacia", también denominados "estudios del mundo real" o "ensayos pragmáticos" han ganado una importancia creciente en el contexto de la medicina basada en la evidencia. Estos estudios siguen estándares metodológicos menos restrictivos que los estudios de fase III en términos de la selección de pacientes, la comedicación y otros temas del diseño, y por lo tanto sus resultados deben ser más generalizables que los de los ensayos de fase III. Los estudios de eficacia, como otros tipos de estudios de fase IV, pueden por lo tanto contribuir al conocimiento de los medicamentos y aportar información relevante además de la que se obtiene de los ensayos de fase III. Sin embargo, el diseño menos restrictivo y los problemas metodológicos inherentes a los estudios de fase IV tienen que ser considerados cuidadosamente. Por ejemplo, la mayor varianza causada por los diferentes tipos de confundentes así como los temas de diseños problemáticos, tales como los criterios para los resultados primarios indiferentes, las condiciones de tratamientos no ciegos, la inclusión de pacientes crónicos refractarios, etc. pueden llevar a conclusiones erróneas. Debido a estos problemas metodológicos, los estudios de eficacia se encuentran principalmente en un nivel de evidencia más bajo, agregando sólo una visión complementaria a los resultados de los estudios de fase III sin desmentir sus resultados.

Avantages et inconvénients des études d'efficacité

Ces dernières années, les « études d'efficacité », aussi appelées « études en conditions réelles » ou « essais pragmatiques » ont acquis une importance croissante dans le contexte de la médecine basée sur les preuves. Ces études suivent des standards méthodologiques moins restrictifs que les études de phase 3 en termes de sélection des patients, de traitement concomitant et d'autres problèmes de conception ; leurs résultats peuvent donc être plus facilement généralisés que ceux des études de phase 3. Les études d'efficacité, comme d'autres types d'études de phase 4, peuvent donc contribuer à la connaissance des traitements et fournir une information pertinente, s'ajoutant à celle des études de phase 3. Il faut cependant soigneusement prendre en compte leur schéma moins restrictif, et les problèmes méthodologiques inhérents aux études de phase 4. Par exemple, une plus grande variance due à différentes sortes de variables confondantes et à des questions délicates de conception, comme des critères de jugement primaires non sensibles, des traitements qui n'ont pas été faits en aveugle, une inclusion de patients chroniques réfractaires etc… peuvent conduire à des conclusions erronées. Les études d'efficacité, du fait de ces problèmes méthodologiques, sont d'un niveau de preuve nettement plus bas, n'apportant qu'un regard complémentaire sur les résultats des études de phase 3, sans les falsifier.

24. Möller HJ. 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*. 2008;258:451-455.

25. Buckley PF. Receptor-binding profiles of antipsychotics: clinical strategies when switching between agents. *J Clin Psychiatry*. 2007;68(suppl 6):5-9..

26. Weiden PJ. Discontinuing and switching antipsychotic medications: understanding the CATIE schizophrenia trial. *J Clin Psychiatry*. 2007;68(suppl 1):12-9.

27. Weiden PJ. Switching antipsychotics: an updated review with a focus on quetiapine. *J Psychopharmacol*. 2006;20:104-118.

28. Naber D. Subjective effects of antipsychotic treatment. *Acta Psychiatr Scand*. 2005;111:81-83.

29. Karow A, Naber D. Subjective well-being and quality of life under atypical antipsychotic treatment. *Psychopharmacology (Berl).* 2002;162:3-10.

30. Lambert M, Naber D. Current issues in schizophrenia: overview of patient acceptability, functioning capacity and quality of life. *CNS Drugs*. 2004;18(suppl 2):5-17;discussion 41-43.

31. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.

32. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey Manual and Interpretation Guide*. Boston, Mass: The Health Institute, New England Medical Center; 1993.

33. Pukrop R, Schlaak V, Möller-Leimkühler AM, et al. Reliability and validity of quality of life assessed by the Short-Form 36 and the Modular System of Quality of Life in patients with schizophrenia and patients with depression. *Psych Res*. 2003;119:63-79.

34. Pukrop R, Möller HJ, Steinmeyer EM. Quality of life in psychiatry: a systematic contribution to construct validation and the development of the integrative assessment tool "modular system for quality of life". *Eur Arch Psychiatry Clin Neurosci*. 2000;250:120-132.

35. Cramer JA, Rosenheck R, Xu W, Thomas J, Henderson W, Charney DS. Quality of life in schizophrenia: a comparison of instruments. Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. *Schizophr Bull*. 2000;26:659-666.

36. Möller HJ. Rating depressed patients: observer- vs self-assessment. *Eur Psychiatry*. 2000;15:160-172.

37. Möller HJ. Standardised rating scales in psychiatry: Methodological basis, possibilities, limitations and descriptions of important rating scales. *World J Biol Psychiatry*. 2009;10:6-26.

38. Paykel ES, Norton KRW. Self-report and clinical interview in the assessment of depression. In: Sartorius N, Ban TA, eds. *Assessment of Depression*. New York, NY: Springer, 1986:356-366.

39. Phillips GA, Van Brunt DL, Roychowdhury SM, Xu W, Naber D. The relationship between quality of life and clinical efficacy from a randomized trial comparing olanzapine and ziprasidone. *J Clin Psychiatry*. 2006;67:1397-1403.

40. Tunis SL, Johnstone BM, Gibson PJ, Loosbrock DL, Dulisse BK. Changes in perceived health and functioning as a cost-effectiveness measure for olanzapine versus haloperidol treatment of schizophrenia. *J Clin Psychiatry*. 1999;60 (suppl 19):38-45;discussion 46:38-45.
41. Franz M, Lis S, Pluddemann K, Gallhofer B. Conventional versus atypical neuroleptics: subjective quality of life in schizophrenic patients. *Br J Psychiatry*. 1997;170:422-425.
42. Awad AG, Voruganti LN. New antipsychotics, compliance, quality of life, and subjective tolerability--are patients better off? *Can J Psychiatry*. 2004;49:297-302.
43. Pyne JM, Sullivan G, Kaplan R, Williams DK. Comparing the sensitivity of generic effectiveness measures with symptom improvement in persons with schizophrenia. *Med Care*. 2003;41:208-217.
44. Rosenheck R, Perlick D, Bingham S, et al. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. *JAMA*. 2003;290:2693-2702.
45. Hofer A, Hummer M, Huber R, Kurz M, Walch T, Fleischhacker WW. Selection bias in clinical trials with antipsychotics. *J Clin Psychopharmacol*. 2000;20:699-702.
46. Meltzer HY, Bobo WV. Interpreting the efficacy findings in the CATIE study: what clinicians should know. *CNS Spectr*. 2006;11(7 Suppl 7):14-24.
47. Möller HJ. Antidepressants: controversies about their efficacy in depression, their effect on suicidality and their place in a complex psychiatric treatment approach. *World J Biol Psychiatry*. 2009;10:180-195.
48. Möller HJ. Are the new antipsychotics no better than the classical neuroleptics? The problematic answer from the CATIE study. *Eur Arch Psychiatry Clin Neurosci*. 2005;255:371-372.
49. Kasper S, Winkler D. Addressing the limitations of the CATIE study. *World J Biol Psychiatry*. 2006;7:126-127.
50. Tandon R, Möller HJ, Belmaker RH, et al. World Psychiatric Association Pharmacopsychiatry Section Statement on Comparative Effectiveness of Antipsychotics in the Treatment of Schizophrenia. *Schizophr Res*. 2007;2008:100:20-38.
51. Adli M, Bauer M, Rush AJ. Algorithms and collaborative-care systems for depression: are they effective and why? A systematic review. *Biol Psychiatry*. 2006;59:1029-1038.
52. Bauer M, Pfennig A, Linden M, Smolka MN, Neu P, Adli M. Efficacy of an algorithm-guided treatment compared with treatment as usual: a randomized, controlled study of inpatients with depression. *J Clin Psychopharmacol*. 2009;29:327-333.
53. Rush AJ. STAR*D: what have we learned? *Am J Psychiatry*. 2007;164:201-204.
54. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163:28-40.
55. Seemüller F, Möller HJ, Obermeier M, et al. Do efficacy and effectiveness samples differ in antidepressant treatment outcome? An analysis of eligibility criteria in randomized controlled trials. *J Clin Psychiatry*. 2010;71:1426-1433.
56. Zimmermann M, Chelminski I, Posternak MA. Generlisability of antidepressants efficacy trials: differences between depressed psychiatric outpatiens who would or would not qualify for an efficacy trial. *Am J Psychiatry*. 2005;162:1370-1372.