Eur Arch Psychiatry Clin Neurosci (2005) 255 : 371–372 DOI 10.1007/s00406-005-0634-2

EDITORIAL

Hans-Jürgen Möller

Are the new antipsychotics no better than the classical neuroleptics?

The problematic answer from the CATIE study

The recently published CATIE study ["Clinical antipsychotic trials of intervention effectiveness", Lieberman J, et al. (September 22, 2005) N Engl J Med 353:12], which was performed in the USA, only showed superiority to the classical neuroleptic perphenazine for the atypical neuroleptic olanzapine, but not for the other atypical neuroleptics (risperidone, quetiapine and ziprasidone).

The results of this study have been dealt with very intensively by the general press, both in Germany and in most other parts of the world. It is understandable that the reports in the general media do not appreciate the methodological difficulties and pitfalls of this study, but reach the relatively across-the-board conclusion that the new drugs for schizophrenia are not superior to those that have been tried and proven for 50 years, and on top of everything are much more expensive. Particularly the way in which the overall problem has been simplified in the general press has resulted in a great deal of uncertainty for the patients and doctors concerned.

It is also to be feared that the health insurance companies and other institutions that are involved in the health system will draw false conclusions from this study, without considering its detailed methodological problems. This could result in a restriction of the prescription or refund possibilities for the new/atypical antipsychotics, although important national and international guidelines on the drug treatment of schizophrenic psychoses present explicitly the leading position of the newer drugs, which has been demonstrated in numerous controlled clinical studies. The new/atypical neuroleptics not only have the advantage that, compared to the classical/typical neuroleptics, they are almost completely free of unwanted extrapyramidal-motor side-effects (this is the central definition criterion of the atypical neuroleptics), but also that they have a broader efficacy spectrum in several symptom areas (such as negative symptoms, depressive symptoms, cognitive disorders as part of schizophrenic psychoses) that are only minimally influenced by the classical neuroleptics.

In order to counteract possible consequences in health policies resulting from the broad discussion of the CATIE study in the general public, it is important to discuss the study under consideration of the methodological problems, and to put the results into perspective. With respect to the methodological errors of this study, one will then reach the conclusion that such an elaborate and expensive study produces more questions than answers.

There is no doubt that the CATIE study is an important study when one considers, for example, the large sample size, complex design with several parallel treatment arms, and the 18-month duration of treatment. However, it is of particular relevance due to the public interest it has already caused. The primary author is without doubt an especially well-known researcher in the area of schizophrenic disorders and antipsychotic treatment. The fact that the study has a public sponsor, the NIMH, and surely cost an enormous amount of money further underlines that it could be expected right from the start to be highly relevant.

One is, therefore, all the more disappointed about the very global and on top of everything very questionable results of the study. The study was primarily planned as a so-called 'effectiveness' study, i. e. a study that analyses whether drugs demonstrate their worth in everyday clinical practice, beyond the selective framework of classical phase III studies. It can be questioned from the outset whether the chosen primary outcome criterion 'discontinuation' is a suitable measurement. This 'effectiveness measure' includes efficacy in the narrower sense and safety, whereby different aspects of tolerability and safety can have different consequences, depending on the detailed characteristic of the tolerability problems. Thus, one can presume a priori that severe extrapyramidal symptoms that occur right at the start result in an early drop-out, the slow development of weight gain rather leads to a later drop-out, and a tardive dyskinesia or metabolic disorder without appreciable weight gain results in a drop-out only years later. This results in the problem that such a rough measurement like 'discontinuation' (or the measurement 'time to discontinuation') causes a biased distortion with respect to the individual neuroleptics being evaluated.

Aside from these detailed problems related to the

chosen primary main outcome criterion and the bias they caused with respect to the individual neuroleptics examined, it should be fundamentally emphasised that not drop-out rates ('discontinuation') or their avoidance are the main focus of treatment of patients with a schizophrenic disorder, but that the aim of treatment is an improvement of symptoms, including cognitive disorders, and quality of life, while avoiding motor side-effects, particularly tardive dyskinesia, which are less common under atypical neuroleptics. For the doctor and patient, the individual adjustment of treatment is much more important than the global question whether one drug is more effective and/or better tolerated than another. In order to adjust treatment individually, it is necessary that several drugs are available, from which the optimal treatment can be chosen for the individual patient.

It should also be considered that it was not an acute treatment study which was then continued as a maintenance study, but that, at the time of recruitment, the patients were rather in a stable psychopathological condition, so that the study was primarily a kind of maintenance study. The average time since the last acute psychotic episode was 3 months. It is much more difficult to ascertain differences between neuroleptics under maintenance conditions than under acute treatment conditions. A less effective drug can, thus, demonstrate 'good efficacy' for a certain time, although we know that, even when stable and particularly remitted patients are switched from a neuroleptic to placebo, it usually takes months until relapses occur.

The extremely high drop-out rate is surprising. It may be due to poor general care conditions, although inadequate drug changeover practices may also play a part. It is noteworthy that a large number of patients discontinued treatment in the first few months of the study. The relevance of this large proportion of early drop-outs may differ for the individual neuroleptics under evaluation. For example, there are generally fewer complications involved in changing from a non-sedating neuroleptic to a sedating one than vice versa.

It is incompatible with a fair study design that one of the drugs (olanzapine) was allowed to be administered at a daily dose of up to 30 mg, which is significantly more than the licensed and recommended daily dose, while all other neuroleptics were applied within the range of the licensed dose, although perphenazine was obviously allocated a rather lower dosage range than is usual. The higher dosing of olanzapine (up to 30 mg/day), which resulted in an average daily dose of 20 mg, may explain the better 'effectiveness' (in the sense of a lower drop-out rate due to better efficacy) of olanzapine. The fact that, even at this high dosage, olanzapine still had a fairly favourable drop-out rate (despite weight gain and changes of metabolic parameters) shows that this neuroleptic obviously has a high general tolerability margin, and that adverse events such as weight gain only result in drop-out after a considerable delay.

The decisive problem is that after randomised allocation of the patients to the various treatment groups, all patients in the perphenazine group who presented with tardive dyskinesia at the start of the study were excluded. As can be seen from the comparison of the numbers of patients in the various treatment arms, this amounted to a relatively large number of patients (approximately 80). Such a procedure, which favours one treatment arm and thus automatically discriminates against the others, is, of course, a severe methodological error. It is difficult to understand why this shortcoming in the study design was overseen or even knowingly accepted. The possible explanatory argument that one cannot treat patients with tardive dyskinesia with a classical neuroleptic (and thus harm them further…) does not solve the methodological dilemma as one, therefore, implies in circular argumentation that one actually already knows that perphenazine (similar to other classical neuroleptics) bears a higher risk than neuroleptics of the second generation of causing extrapyramidal effects and thus also tardive dyskinesia. However, the reason for choosing perphenazine from the group of classical neuroleptics was that it may bear less risk in this respect, and that, within the group of classical neuroleptics, it may actually have a rather more favourable profile that may be similar to that of the 'atypicals'. It is known that tardive dyskinesia correlates very strongly with the general risk of extrapyramidal symptoms and that the occurrence of early dyskinesia or parkinsonism is a very good predictor of the later development of tardive dyskinesia. If one removes, in a randomised, control-group study, these risk patients from the treatment arm receiving a typical neuroleptic, but leaves them in the groups receiving an atypical neuroleptic, such a biased design has the consequence that one can no longer discern the actual advantages of the atypicals compared to the typicals. As is known, the better extrapyramidal tolerability is one of the main advantages of the 'atypical neuroleptics'. It is, therefore, almost contrary to the expectations that, even so, the study found a certain advantage in this respect for atypicals compared to perphenazine. Of course, perphenazine itself is completely underrated due to the methodological shortcomings described above.

The study has unfortunately received much too much publicity, particularly in the general press, where it is portrayed as showing that, for the most part, neuroleptics of the second generation are not better (but much more expensive!) than neuroleptics of the first generation. It is not possible to reach this conclusion because of principal methodological failings, as described above. The chance to verify in a broadly applied 'effectiveness' study the advantages of atypicals, which are known from numerous double-blind, randomised phase III studies, compared to a representative from the group of classical neuroleptics was, unfortunately, wasted due to severe methodological shortcomings.

Prof. Hans-Jürgen Möller, Munich