

Defining 'Response' in Antipsychotic Drug Trials: Recommendations for the Use of Scale-Derived Cutoffs

Stefan Leucht^{*1}, John M Davis², Rolf R Engel³, John M Kane⁴ and Stefan Wagenpfeil⁵

¹Department of Psychiatry and Psychotherapy, Technische Universität München, Klinikum rechts der Isar, Munich, Germany; ²Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA; ³Department of Psychiatry, Psychiatrische Klinik der Ludwig-Maximilian-Universität München, München, Germany; ⁴Department of Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, NY, USA; ⁵Institute of Medical Statistics and Epidemiology of the Technische Universität München, München, Germany

Scale-derived cutoff points are frequently used to define 'response' in antipsychotic drug trials. This procedure is useful, because responder rates can be understood more intuitively than a difference in means of rating scales. As various definitions of response have been used, we examined original participant data to assess whether the choice of the Brief Psychiatric Rating Scale-based response cutoff had an impact on the results of seven ($n = 1870$) antipsychotic drug trials in schizophrenia. We also analyzed whether the chronicity of the illness has an impact on the question of which cutoff is most sensitive in detecting differences between drugs. The results in terms of p -values and response rate differences varied substantially in dependence on the cutoff chosen. The use of response rate ratios attenuated the variability. In contrast to a widely held belief, low response cutoffs were not more sensitive in detecting differences between drugs than higher cutoffs. In more chronic, less responsive participants, there was a trend for higher cutoffs to be less sensitive in detecting differences between drugs than lower ones. The results of clinical trials depend considerably on the response cutoff chosen. Therefore, the cutoff should never be chosen *post hoc*, a large range of cutoffs should be presented and the *a priori* choice of the primary cutoff should be based on clinical relevance. The use of ratios rather than differences attenuates the variability. Cutoffs need to be calculated on the basis of 0–6 rather than on 1–7 scoring systems. We suggest a table presenting responder rates in 25 percent steps covering the whole range up to 100% reduction from baseline, which could be displayed together with recently presented criteria for remission.

Neuropsychopharmacology (2007) 32, 1903–1910; doi:10.1038/sj.npp.1301325; published online 7 February 2007

Keywords: response; cutoff; remission; outcome; schizophrenia; antipsychotic

INTRODUCTION

In many diseases in which 'hard' outcomes such as death are not appropriate measures of efficacy, rating scales are used for the evaluation of treatments. Schizophrenia is such a disease and the Brief Psychiatric Rating Scale (BPRS; Overall and Gorham, 1962) is one of the most frequently used instruments for evaluating its psychopathology. To evaluate the effects of different interventions, the mean BPRS end point score or the mean change from baseline score has usually been used to examine differences between interventions in clinical trials. Although such analyses are usually the primary outcomes in antipsychotic drug trials, one problem is that the translation of the results into

clinical practice is difficult. We do not know well what, for example, a difference which is highly statistically significant but amounts to only two BPRS points in magnitude means from a clinical perspective. Therefore, in clinical trials BPRS-defined response rates are usually analyzed as additional secondary outcomes. Knowing how many participants have responded to treatment can be understood more intuitively by clinicians than an x -point difference on a rating scale. However, a number of problems arise from this procedure: one is that a number of different cutoffs have been used in clinical studies, for example, at least 20% (Kane *et al*, 1988; Marder and Meibach, 1994), 30% (Small *et al*, 1997; Arvanitis *et al*, 1997), 40% (Beasley *et al*, 1996), or 50% (Peuskens and Link, 1997) of the baseline score, so that there is no consensus as to which is the most appropriate one. If a cutoff is chosen *post hoc*, this is potentially an inappropriate manipulation of the data. In recent trials, an at least 20% cutoff has often been used, although recent analyses showed that this level of response does not even mean minimal improvement according to the Clinical Global Impression (CGI) of the raters (Leucht *et al*,

*Correspondence: PD Dr S Leucht, Department of Psychiatry and Psychotherapy, Technische Universität München, Klinikum rechts der Isar, Ismaningerstr. 22, 81675 Munich, Germany, Tel: +49 89 4140 4249, Fax: +49 89 4140 4888, E-mail: Stefan.Leucht@lrz.tum.de
Received 7 June 2006; revised 24 October 2006; accepted 7 December 2006

2005a, b, 2006). One reason for this development may be an assumption that using a low cutoff increases the ability to find statistically significant differences between drugs. A further problem is that information is lost by dichotomizing the results of the BPRS using a cutoff, so that the results may vary depending on the cutoff used. To the best of our knowledge, neither of the latter two issues has been systematically examined. We therefore reanalyzed a number of antipsychotic drug trials concentrating on three questions:

1. To what extent do the results of clinical trials differ in dependence on the response cutoff applied?
2. Are there differences in the variability of different statistical measures?
3. Does the chronicity of the participant's illness have an impact on the sensitivity of the cutoff in showing between-drug differences?
4. Finally, we suggest a simple table that presents responder rates in a comprehensive way.

PATIENTS AND METHODS

The Database

Original participant data from seven trials (*n* at baseline = 1870) that compared amisulpride with other antipsychotics were reanalyzed *post hoc* (see Table 1). All

studies were randomized and all but one (Colonna *et al*, 2000) were double-blind. Six trials used the original BPRS (Overall and Gorham, 1962), and for one of them (Sèchter *et al*, 2002) the BPRS was derived from the Positive and Negative Syndrome Scale (PANSS) (Kay *et al*, 1987). Each trial included participants with schizophrenia or schizophreniform disorder according to DSM-III-R or DSM-IV (American Psychiatric Association, 1987, 1994). Four studies compared amisulpride with haloperidol (Möller *et al*, 1997; Carrière *et al*, 2000; Puech *et al*, 1998; Colonna *et al*, 2000), one with flupenthixol (Wetzel *et al*, 1998), and two with risperidone (Sèchter *et al*, 2002; Peuskens *et al*, 1999). With one exception (Carrière *et al*, 2000), all studies required various minimum scores as eligibility criteria to assure that the participants had florid positive symptoms. The participants in the study without scale-derived minimum scores (Carrière *et al*, 2000) were all in-patients and had a mean BPRS of 65 at baseline, so that highly symptomatic participants were also involved in this study. Sixty-one participants of one fixed-dose study (Puech *et al*, 1998) who had received a potentially subtherapeutic 100 mg/day amisulpride dose were excluded. The other three amisulpride doses of the latter study were pooled in one group in our analyses. The mean BPRS total score of the remaining 1809 participants was 57.8 ± 11.7 (mean \pm SD) at baseline. There were 1119 men and 690 women: age 36.1 ± 10.9 years, weight 70.8 ± 14.6 kg, height 170 ± 9 cm. Our data set includes all randomized controlled amisulpride trials in participants with schizophrenia and positive

Table 1 Characteristics of the Included Studies

Study	Antipsychotic drugs and daily dose (mg)	<i>n</i>	Weeks	Mean BPRS at baseline	Selected eligibility criteria
Möller <i>et al</i> , 1997	AMI 800/600 ^a	95	6	61	In-patients with paranoid, disorganized, or undifferentiated schizophrenia BPRS psychotic subscore ≥ 12 and at least 2 psychotic items ≥ 4
	HAL 20/15 ^a	96			
Wetzel <i>et al</i> , 1998	AMI 1000/600 ^a	70	6	53	Acutely admitted in-patients with paranoid or undifferentiated schizophrenia BPRS total score ≥ 36 , but no predominant negative symptoms defined as SANS composite score > 55
	FLU 25/15 ^a	62			
Puech <i>et al</i> , 1998	AMI (100 ^b ; 400;800;1200) HAL 16	(61;64;65;65) 64	4	61	In-patients with acute exacerbations of paranoid, disorganized, or undifferentiated schizophrenia BPRS psychotic subscore ≥ 12 and at least 2 psychotic items ≥ 4
Colonna <i>et al</i> , 2000	AMI 200–800 HAL 5–20	370 118	51	56	In-patients or outpatients with acute exacerbations of paranoid, disorganized, or undifferentiated schizophrenia BPRS psychotic subscore ≥ 12 and at least 2 psychotic items ≥ 4
Carrière <i>et al</i> , 2000	AMI 400–1200 HAL 10–30	94 105	17	65	In-patients with paranoid schizophrenia or schizophreniform disorder
Peuskens <i>et al</i> , 1999	AMI 800 RIS 8	115 113	8	55	In-patients or outpatients with paranoid, disorganized, or undifferentiated schizophrenia BPRS total score ≥ 36 , BPRS psychotic subscore ≥ 12 and at least 2 psychotic items ≥ 4
Sèchter <i>et al</i> , 2002	AMI 400–1000 RIS 4–10	152 158	51 ^c	54	In-patients or outpatients with schizophrenia PANSS total score 60–120, no predominant negative symptoms defined as 3 or more PANSS negative items ≥ 4

n = number of participants; mg = milligram; AMI = amisulpride; HAL = haloperidol; FLU = flupenthixol; RIS = risperidone; BPRS = Brief Psychiatric Rating Scale; SANS = Scale for the Assessment of Negative Symptoms; PANSS = Positive and Negative Syndrome Scale.

^aAll participants were started at the higher dose, which could then be reduced.

^bThis potentially subtherapeutic dose group was excluded.

^cIn the original report, only the results at 6 months were reported, but there was a double-blind extension to a total of 12 months that we used for our analysis.

symptoms available in the manufacturer's database with the exception of one study that was not available when the project was begun (Mortimer *et al*, 2004).

Statistical Analysis

We calculated the percentage BPRS change from baseline to study end point (B%) for each participant using the formula: $B\% = (B_0 - B_{LOCF}) \times 100 / (B_0 - 18)$, where B_0 = BPRS at baseline, B_{LOCF} = BPRS at last observation, and with 18 being the minimum score of the BPRS in the 1–7 rating system. The data were analyzed using a last-observation-carried-forward method, where missing cases were replaced by the last available observation, even if a participant dropped out before the first postbaseline rating (this was the case for only 24 participants so that our strict 'once randomized—analyzed' approach that has also been applied in reviews of the Cochrane Schizophrenia Group (Adams *et al*, 2005) should not have yielded results differing, for example, from those including only those participants with at least one postbaseline assessment). We then analyzed whether there was a difference in responder rates between amisulpride and comparator drugs. Ten different percentage BPRS cutoffs were used to define response: at least 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% reduction of the baseline score. Although cutoffs above 60% are usually not 'realistic' as primary thresholds in recent trials that include many refractory patients, they were necessary for drawing our curves. They may also be useful in illustrating the distribution of the data (see Discussion).

In order to examine to what extent the results based on the various cutoffs differed, the following three measures were calculated for each cutoff:

(1) *P*-values for the difference between amisulpride and comparator drugs using Fisher's exact test. Two-sided *p*-values were analyzed as a measure related to type I error. (2) The response rate difference (RRD) and (3) the response rate ratio (RRR) between amisulpride and the comparator antipsychotics. RRD was calculated by the formula $\% \text{responders}_A - \% \text{responders}_C$, and RRR as $\% \text{responders}_A / \% \text{responders}_C$ with $\% \text{responders}_A$ being the percentage number of participants who responded in the amisulpride group and $\% \text{responders}_C$ being the like percentage in the control group. Although the *p*-value is a conditional probability and thus a measure related to statistical probability, RRD and RRR are measures of the magnitude of the difference between amisulpride and comparators (ie they are effect sizes for dichotomous data). Please note that the number needed to treat (NNT) is derived from the RRD. It is simply the inverse of the RRD ($\text{NNT} = 1/\text{RRD}$). Thus, our results on RRD also reflect the variability of NNT. As NNT is a clinically intuitive measure indicating how many participants need to be treated with (here) amisulpride to have one more responder, we indicate it in Figure 2 in addition to the RRD.

We also calculated between-group effect sizes expressed as Cohen's *d* for the mean percentage BPRS change from baseline. As our paper is on dichotomous response cutoffs rather than on using the BPRS as a dimensional scale, these effect sizes are briefly shown as a footnote to Figure 1 for reference.

We analyzed the single studies separately and all studies combined. Finally, to test whether chronicity of the illness

had an effect on the question of which cutoff showed the highest difference between amisulpride and comparators, the participants were divided into two groups according to the median of their duration of illness. Participants who are more chronic should on average show less improvement of symptoms in terms of percentage BPRS reduction than those who are less chronic. The hypothesis was that relatively lower cutoffs should show more pronounced differences between amisulpride and other antipsychotics in the more chronic participants than in the less chronic group. One study (Wetzel *et al*, 1998) had to be excluded from this analysis because information on the participants' duration of illness was not recorded. The analyses were made using SPSS version 11.5 for Windows, Excel for Windows 2000, and StatXact version 5.

RESULTS

Overall Response in the Trials

The mean change in the BPRS total score of all participants ($n = 1809$) combined was 18.3 ± 16.6 (median 19.0). The mean percentage BPRS change of all studies was $45.4 \pm 39.3\%$ (median 53.6) with a range of 41.8–55.3% considering the single studies. This shows that the participants included in the studies responded comparatively well on the average. For example, the mean percentage reduction of the PANSS in two pivotal studies comparing risperidone with haloperidol was less than 20% (Davis and Chen, 2002). With the exception of Sèchter *et al* (2002) in which the efficacy of both drugs was similar, there was at least a trend in favor of amisulpride in all studies. The mean percentage BPRS change at end point (LOCF), and the response rates based on four degrees of percentage BPRS change (<25, 25–49, 50–74, $\geq 75\%$) in the single studies can be looked up on the journal's website (see Supplementary Table 2).

Statistical Measures Resulting from the Use of Different Cutoffs

Figure 1 displays the *p*-values derived from Fisher's exact tests for the different BPRS percentage cutoffs. The results of the single studies are shown as well as the pooled results of all 'amisulpride vs other antipsychotics' studies. The main result illustrated by the figure is that the *p*-values differed substantially in dependence on the % BPRS cutoff used. Sometimes the *p*-values changed quite abruptly from one cutoff to the next higher one. For example, in Möller *et al* (1997) the *p*-value of the cutoff 'at least 40% BPRS reduction' was only 0.24, whereas the cutoff 'at least 50% BPRS reduction' showed a significant difference between groups with $p = 0.03$. Likewise, in Puech *et al* (1998) the *p*-value at the 'at least 30%' cutoff was 0.0095, whereas at the 'at least 50%' cutoff it was 0.37.

Figure 2 showing the absolute RRD is consistent with the heterogeneity found for the *p*-values. For example, in Möller *et al* (1997) the absolute RRD at the 60% cutoff was 19%, whereas at the 40% cutoff it was only 9%. Or in Sèchter *et al* (2002) the RRD at the 50% cutoff was 6%, whereas at the 60% cutoff it was –1% in favor of risperidone.

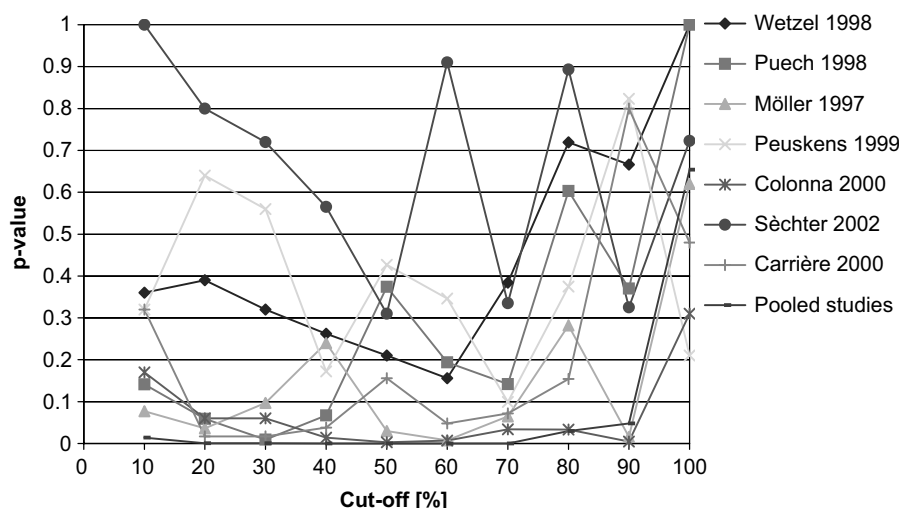


Figure 1 P-values derived from Fisher's exact tests obtained for different cutoffs. The between-group effect sizes expressed as Cohen's d 's were: pooled studies $d = 0.18$, $p < 0.001$ (Wetzel et al, 1998); $d = 0.25$, $p = 0.16$ (Puech et al, 1998); $d = 0.22$, $p = 0.07$ (Möller et al, 1997); $d = 0.27$, $p = 0.07$ (Peuskens et al, 1999); $d = 0.15$, $p = 0.25$ (Colonna et al, 2000); $d = 0.24$, $p = 0.01$ (Sèchter et al, 2002); $d = -0.009$, $p = 0.94$ (Carrière et al, 2000); $d = 0.35$, $p = 0.01$.

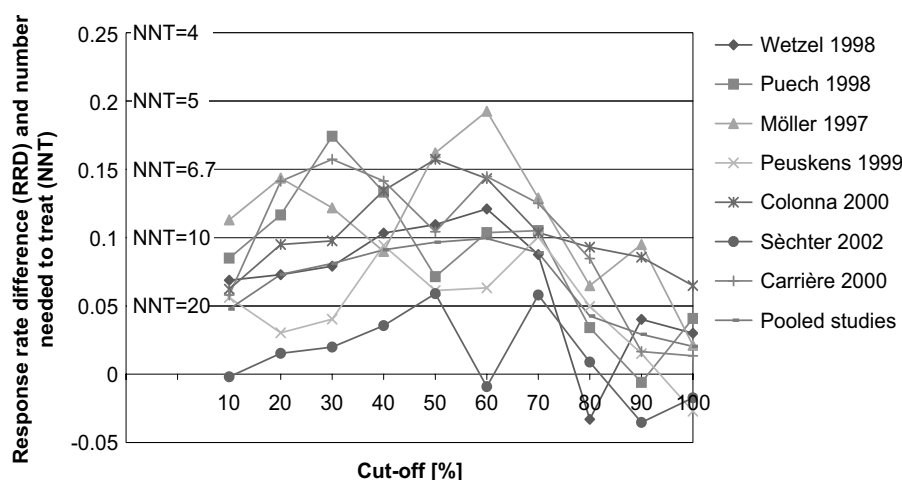


Figure 2 RRDs and NNT obtained for different cutoffs.

A different picture was found when the RRRs were analyzed in Figure 3. Here, the results remained relatively stable approximately up to the cutoff of at least 70% BPRS reduction, whereas above the 70% cutoff in some studies the RRR changed abruptly, probably because of few participants reaching these very high response thresholds in the studies. We did a *post hoc* comparison of the variation coefficients ($V = SD/|mean|$) and their 95% confidence intervals (CI) of RRD and RRR. When the whole range of cutoffs (10–100%) was examined, no difference in variation coefficients (V of RRR 0.67, CI 0.56–0.83; V of RRD 0.75, CI 0.62–0.94) was found, probably because of the extremely variable results of RRR at cutoffs, higher than 70% BPRS reduction. However, when only cutoffs between 10 and 70% were included in the analysis, the variation coefficient of RRR was indeed significantly lower than that of RRD (V of RRR 0.12, CI 0.10–0.14; V of RRD 0.46, CI 0.40–0.56). The difference was statistically significant, because the 95% CI did not overlap.

All three figures illustrate that the two lowest cutoffs (at least 10% and at least 20% BPRS reduction) are not the ones

that showed the largest differences between amisulpride and comparators according to any of the four measures. The cutoffs with the highest sensitivity for showing differences between treatments roughly lay in the range between 30 and 70% rather than below.

Finally, Figure 4 plots the results of the pooled studies when the participants were divided into two groups of greater and lesser chronicity according to the median duration of their illness at baseline. Results based on RRD were used as an example here, but similar figures using p -value and RRR can be found on the journal's website (see Supplementary Figures 5 and 6). The hypothesis was that the more chronic participants are less responsive to treatment, and therefore the cutoffs showing the highest sensitivity to differences between drugs should be lower than in the less chronic participants. Although the figures do not show a dramatic effect in this sense, it appears that there is a certain shift to the right, that is, in the less chronic participants the values indicating greater differences between drugs are reached only at somewhat higher cutoffs than in the more chronic participants. The effect was

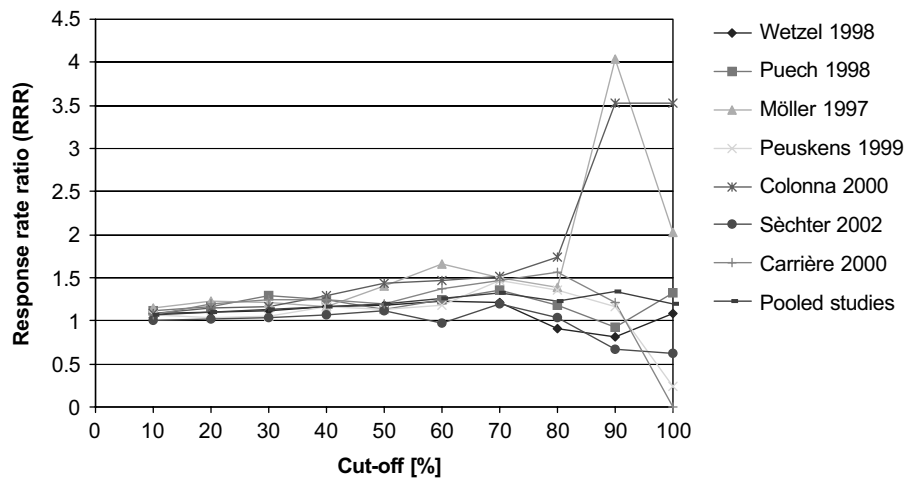


Figure 3 RRRs obtained for different cutoffs.

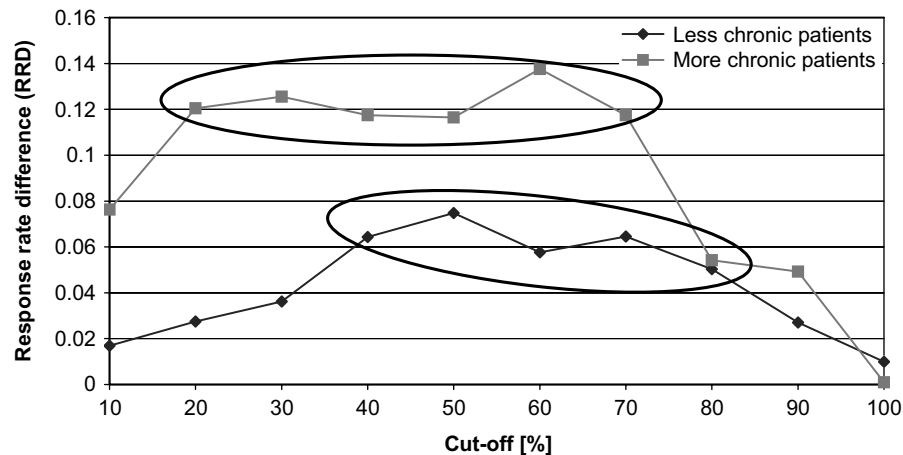


Figure 4 Pooled studies—less chronic vs more chronic participants—RRD. Circles illustrate the cutoffs showing the highest differences between drugs.

illustrated by circling the cutoffs showing the highest differences between drugs. The nature of the difference is purely qualitative, and two factors may explain why the difference was not more pronounced: (1) The difference in the reduction of symptoms compared to baseline between less and more chronic participants was small (mean percentage BPRS reduction 47.8 ± 36.7 vs $41.3 \pm 41.2\%$; median 54.2 vs 50.7%) and (2) The difference between drugs was more pronounced among the more chronic participants than among the less chronic participants (see Figure 4). Both effects may have overshadowed differences between cutoffs among more and less chronic participants. Furthermore, there was no clear chronicity effect using RRR, possibly again because the variability of RRR was reduced by taking the baseline response rate into account (see Supplementary Figures 5 and 6).

DISCUSSION

The results of clinical trials can vary to an important extent if different scale-derived cutoff values are used to define response. If the cutoff is chosen *post hoc*, the results can be

easily biased in favor of one compound. Furthermore, the highest difference between amisulpride and comparator antipsychotics was not found at very low cutoffs, such as, for example, an at least 20% BPRS reduction, but rather at medium high cutoffs in a broad range between approximately 30 and 70% BPRS reduction. In more chronic, less responsive participants, the cutoffs showing the highest sensitivity for detecting between-drug differences were somewhat lower than in less chronic participants. The use of RRRs rather than RRDs reduced the variability of the findings, at least as long as the response rates were well above zero.

It needs to be pointed out that our analysis is not an evaluation of the efficacy of amisulpride. A number of meta-analyses on this question do exist (Davis *et al*, 2003; Mota Neto *et al*, 2004; Leucht *et al*, 2002). We attempted only to evaluate the impact of different cutoffs to define response on a number of basic statistical measures. Therefore, more sophisticated statistical approaches such as non-inferiority tests were not applied. Given the theoretical nature of our analyses, we found it also justified to pool the results of all studies in an analysis ‘amisulpride vs other antipsychotic drugs’ irrespective of whether the comparator was a typical

or an atypical antipsychotic. In an analysis of the efficacy of amisulpride, this procedure might not have been considered meaningful.

Antipsychotic drug trials usually compare the means of rating scales at end point in each group rather than comparing responder rates based on cutoff values as primary outcomes. Thus, researchers and statisticians are aware of the general problem. However, to the best of our knowledge this is the first published attempt to analyze this problem in a systematic manner by looking at a large range of response cutoffs in a number of relatively large studies. Although *p*-values and RRD varied quite substantially depending on the cutoff applied, this variability was reduced when RRRs rather than RRDs were calculated as effect size measures, but only as long as response rates were well above zero. The reason for this phenomenon might be that in contrast to the RRD, the RRR takes the baseline risk into account. For this reason, relative risks rather than absolute risk reductions have recently been recommended as effect size measures for meta-analyses, a recommendation that is supported by our results (Deeks, 2002). However, in reports of clinical trials in psychiatry, relative risks have usually not been presented to date. Rather, results are presented as the percentage of participants, for example, relapsed in the intervention group and in the control group. Relative risks are difficult for inexperienced readers to interpret (Boissel *et al*, 1999). For example, if a clinician were told that the response rate of a drug is 25% higher than that of placebo, he would probably intuitively think something like 50 vs 25% or 75 vs 50% rather than 6 vs 8% (which means a 25% higher RRR, but only a 2% RRD). Thus, intuitivity is reduced by applying relative risks. It has, therefore, been recommended that the choice of risk differences and relative risk reduction should depend on the respective situation (Egger *et al*, 1997). We also find it crucial that instead of merely presenting effect sizes, the raw numbers underlying the effect sizes need to be presented. The presentation of raw numbers (response rate in group A and response rate in group B) facilitates the interpretation of the data.

Furthermore, there has been a trend toward using very low cutoffs such as an at least 20% reduction of the BPRS or the PANSS in recent years. For example, all the pivotal risperidone studies used this 20% cutoff to define response (Marder and Meibach, 1994; Chouinard *et al*, 1993; Peuskens *et al*, 1995; Hoyberg *et al*, 1993; Claus *et al*, 1992). We believe that one tacit assumption of this strategy is that low cutoffs increase the power to find statistically significant differences between interventions. In our data set, however, this was not the case because the cutoffs that were most sensitive to differences between drugs were usually greater than 20%.

This finding may depend on how well the participants generally responded in a trial. The sensitivity analysis splitting the pooled database into two groups according to the participants' chronicity showed that the most sensitive cutoff was somewhat lower in the more chronic participants than in the less chronic participants. Two facts may explain why the difference was not more pronounced: (1) The difference *between drugs* was larger among the more chronic participants than among the less chronic participants (see Figure 4) so that this effect may have over-

shadowed the differences in cutoffs, and (2) the difference in reduction of symptoms compared to baseline *between less and more chronic participants* was small (mean percentage BPRS reduction 47.8 ± 36.7 vs $41.3 \pm 41.2\%$; median 54.2 vs 50.7%). Thus, although in our data set even the more chronic participants responded relatively well, we assume that in even less responsive participants the most sensitive cutoff would have been lower. For example, in two pivotal risperidone studies the average PANSS reduction was less than 20% (Marder and Meibach, 1994; Chouinard *et al*, 1993; Davis and Chen, 2002). If higher cutoffs than at least 20% PANSS reduction had been used, it might have been impossible to find statistically significant differences.

Despite the variability of results in dependence on the cutoffs selected, we believe that the analysis of response rates is still important at least as a secondary outcome, because this procedure has one major advantage: response rates can be understood more intuitively by clinicians than an *x*-point difference of the BPRS total score at study end point. We thus derive the following suggestions from our analysis:

1. The greatest mistake that can be made is to choose the cutoff showing the highest superiority of a new antipsychotic *post hoc*. Our results show that such a procedure would easily allow a manipulation of the results.
2. Our data suggest that a broad range of cutoffs should be presented so that it can be evaluated whether the effects were consistent. However, according to our experience with the literature this is performed very rarely. Authors at best present a number of cutoffs (eg at least 20, 30, 40, and 50%). Rather than presenting only one or a number of arbitrarily chosen cutoffs, we suggest a table that presents BPRS- or PANSS-derived responder rates in steps of 25% (less than 25% reduction from baseline, 25–49% reduction, 50–75% reduction, >75% reduction, see Table 2). The advantage of this table is that it provides an impression of the *distribution of the results* by displaying the whole range of responder rates including the extremes. The example shows that 75–100% BPRS reduction may not always be as rare as generally thought. The table can also be put together easily and does not need a lot of space. It has been used in the Chinese literature (see eg Wang *et al*, 2001). Similar tables could be used when the CGI-improvement score or the CGI-severity score is used to define response (Table 3). They should present the numbers and percentages of participants meeting each of the seven CGI scores.
3. To avoid the problem of multiple testing, statistical tests may be applied not to each cutoff, but rather to one cutoff chosen *a priori*. For this purpose, the presentation of the responder rates in Table 3 easily allows a calculation of the participants with, for example, at least 50% BPRS reduction.
4. The choice of this cutoff might be made according to clinical relevance. In three previous publications, we found that a 25% BPRS/PANSS reduction roughly corresponded to a CGI (Guy, 1976) of 'minimally better', whereas the 50% cutoff roughly corresponded to 'much better' (Leucht *et al*, 2005a,b, 2006). From a clinical

Table 2 Suggestion of a Simple Table for the Presentation of Percentage BPRS/PANSS-Derived Responder Rates

	Total <i>n</i>	<25% PANSS/BPRS reduction <i>n</i> (%)	25–49% PANSS/BPRS reduction <i>n</i> (%)	50–74% PANSS/BPRS reduction <i>n</i> (%)	75–100% PANSS/ BPRS reduction <i>n</i> (%)
Intervention group	1092	278 (25.5)	190 (17.4)	294 (26.99)	330 (30.2)
Control group	716	239 (33.4)	137 (19.1)	173 (24.2)	167 (23.3)

In this example, the intervention group is all participants in the amisulpride groups and the control group is all participants in the comparator antipsychotic groups.

Table 3 Suggestion of a Simple Table for the Presentation of CGI-Derived Responder Rates

CGI-improvement score	Total <i>n</i>	Very much worse	Much worse	Minimally worse	Unchanged	Minimally better	Much better	Very much better
Score		7	6	5	4	3	2	1
Intervention group	<i>n</i>	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Control group	<i>n</i>	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)

CGI-severity score	Total <i>n</i>	Extremely ill	Severely ill	Markedly ill	Moderately ill	Mildly ill	Borderline mentally ill	Normal, not at all ill
Score		7	6	5	4	3	2	1
Intervention group	<i>n</i>	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Control group	<i>n</i>	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)

perspective, we think that in acutely ill, well-responding participants a cutoff reflecting much improvement (50% BPRS/PANSS reduction) would be more meaningful than minimal improvement (25% BPRS/PANSS reduction). The 25% cutoff may still be useful in treatment of refractory participants in which even small changes of psychopathology can be important.

5. A basic problem is that percentage BPRS/PANSS reduction is often not calculated correctly, because the minimum score of 18/30 is not subtracted. Owing to the 1–7 scoring system of both scales each participant always has at least 18/30 points meaning ‘no symptoms’. Thus, a reduction of the PANSS from 100 at baseline to 50 at the end of the study is not a 50% reduction, but rather a 71.4% reduction (the correct calculation is $70 \frac{100-30}{100} = 71.4\%$). Not subtracting the 18/30 minimum means to underestimate percentage BPRS/PANSS reduction. Or, if the 30 minimum points are not subtracted, a 100% PANSS reduction is simply impossible, because every participant will have at least a score of 30 at end point. To some readers, this point may appear trivial, but there are *multiple* examples in the literature where percentage BPRS/PANSS reduction was not calculated correctly, even by professional statisticians from pharmaceutical companies (eg Emsley *et al*, 2006; Tollefson *et al*, 2001). Indeed, there is no rational why a 1–7 scoring system should be better than a 0–6 scoring system (Thompson *et al*, 1994). The decision to rate the PANSS items 1–7 rather than 0–6 was more or less arbitrary (L Opler, personal communication), but makes it even more difficult to understand PANSS total scores intuitively, because the reader must always subtract 30 points which do not relate to any symptoms.

6. Last but not least, ‘remission criteria’ for schizophrenia have recently been presented (Andreasen *et al*, 2005). The advantage of the remission criteria is that they show how many participants were without significant symptoms at the end of a study. In contrast, even an important reduction of the PANSS from baseline does not rule out that the participants were still symptomatic. For example, a change of the PANSS from 140 to 70 is a 50% reduction (0–6 scale), but the participant is still symptomatic. The disadvantage of the remission criteria is that they do not reflect the change of symptoms. In a study with low levels of symptoms at baseline, many participants may be in remission at its end although the overall change of symptoms was small. In many studies, the use of both remission and response criteria may be warranted.

ACKNOWLEDGEMENTS

We are indebted to SanofiAventis for allowing us to analyze individual participant data from their databases. The study was supported by the APA/AstraZeneca Young Minds in Psychiatry Award 2004. The computational assistance of Caterina Riehle and of Gudrun Wagenpfeil is gratefully acknowledged.

REFERENCES

- Adams CE, Coutinho E, Duggan L, Leucht S, Srisurapanont M, Tharyan P, Cochrane Schizophrenia Group (2005). *The Cochrane Library*. John Wiley & Sons Ltd: Chichester, UK.
- American Psychiatric Association (1987). *Diagnostic and Statistical Manual for Mental Disorders*, 3rd revision, revised (DSM-III-R). American Psychiatric Association: Washington, DC.

- American Psychiatric Association (1994). *Diagnostic and Statistical Manual for Mental Disorders*, 4th edn. American Psychiatric Association: Washington, DC.
- Andreasen N, Carpenter W, Kane J, Lasser R, Marder S, Weinberger D (2005). Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry* 62: 441–449.
- Arvanitis LA, Miller BG, Seroquel Trial 13 Study Group (1997). Multiple fixed doses of 'seroquel' (quetiapine) in patients with acute exacerbation of schizophrenia: a comparison with haloperidol and placebo. *Biol Psychiatry* 42: 233–246.
- Beasley CM, Tollefson GD, Tran P, Satterlee W, Sanger T, Hamilton S, Olanzapine HGAD Study Group (1996). Olanzapine versus haloperidol and placebo. Acute phase results of the American double-blind olanzapine trial. *Neuropsychopharmacology* 14: 111–123.
- Boissel JP, Cucherat M, Li W, Chatellier G, Gueyffier G, Buyse M et al (1999). The problem of therapeutic efficacy indices. 3. Comparison of the indices and their use. *Therapie* 54: 405–411.
- Carrière P, Bonhomme D, Lempérière T (2000). Amisulpride has superior benefit:risk profile to haloperidol in schizophrenia: results of a multicentre, double-blind study (the Amisulpride Study Group). *Eur Psychiatry* 15: 321–329.
- Chouinard G, Jones B, Remington G (1993). Canadian placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenic patients. *J Clin Psychopharmacol* 13: 25–40.
- Claus A, Bollen J, De Cuyper H (1992). Risperidone versus haloperidol in the treatment of chronic schizophrenic inpatients: a multicentre double-blind comparative study. *Acta Psychiatr Scand* 85: 295–305.
- Colonna L, Saleem P, Dondey-Nouvel L, Rein W, Amisulpride Study Group (2000). Long-term safety and efficacy of amisulpride in subchronic or chronic schizophrenia. *Int Clin Psychopharmacol* 15: 13–22.
- Davis JM, Chen N (2002). Clinical profile of an atypical antipsychotic: risperidone. *Schizophr Bull* 28: 43–61.
- Davis JM, Chen N, Glick ID (2003). A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry* 60: 553–564.
- Deeks JJ (2002). Issues in the selection of a summary statistic for meta-analyses of clinical trials with binary outcomes. *Stat Med* 21: 1575–1600.
- Egger M, Davey Smith G, Philipps AN (1997). Meta-analysis: principles and procedures. *BMJ* 315: 1533–1537.
- Emsley R, Rabinowitz J, Medori R (2006). Time course for antipsychotic treatment response in first-episode schizophrenia. *Am J Psychiatry* 163: 743–745.
- Guy W (1976). Clinical Global Impressions. *ECDEU Assessment Manual for Psychopharmacology*, revised (DHEW Publ No ADM 76-338) National Institute of Mental Health: Rockville, MD. pp 218–222.
- Hoyberg OJ, Fensbo C, Remvig J, Lingjaerde O, Sloth-Nielsen M, Salvesen I (1993). Risperidone versus perphenazine in the treatment of chronic schizophrenic patients with acute exacerbations. *Acta Psychiatr Scand* 88: 395–402.
- Kane JM, Honigfeld G, Singer J, Meltzer H, Clozaril Collaborative Study Group (1988). Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch Gen Psychiatry* 45: 789–796.
- Kay SR, Fiszbein A, Opler LA (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13: 261–275.
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR (2005a). Clinical implications of BPRS scores. *Br J Psychiatry* 187: 363–371.
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR (2005b). What does the PANSS mean? *Schizophr Res* 79: 231–238.
- Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR (2006). Linking the PANSS, BPRS and CGI: clinical implications. *Neuropsychopharmacology* 31: 2318–2325.
- Leucht S, Pitschel-Walz G, Engel R, Kissling W (2002). Amisulpride—an unusual atypical antipsychotic. A meta-analysis of randomized controlled trials. *Am J Psychiatry* 159: 180–190.
- Marder SR, Meibach RC (1994). Risperidone in the treatment of schizophrenia. *Am J Psychiatry* 151: 825–835.
- Möller HJ, Boyer P, Fleurot O, Rein W (1997). Improvement of acute exacerbations of schizophrenia with amisulpride: a comparison with haloperidol. *Psychopharmacology* 132: 396–401.
- Mortimer A, Martin S, Loo H, Peuskens J (2004). A double-blind, randomized comparative trial of amisulpride versus olanzapine for 6 months in the treatment of schizophrenia. *Int Clin Psychopharmacol* 19: 63–69.
- Mota Neto JIS, Lima MS, Soares BGO (2004). Amisulpride for schizophrenia (Cochrane Review). *The Cochrane Library*. John Wiley & Sons Ltd: Chichester, UK.
- Overall JE, Gorham DR (1962). The Brief Psychiatric Rating Scale. *Psychol Rep* 10: 790–812.
- Peuskens J, Bech P, Möller HJ, Bale R, Fleurot O, Rein W, Amisulpride Study Group (1999). Amisulpride vs risperidone in the treatment of acute exacerbations of schizophrenia. *Psychiatry Res* 88: 107–117.
- Peuskens J, Link CGG (1997). A comparison of quetiapine and chlorpromazine in the treatment of schizophrenia. *Acta Psychiatr Scand* 96: 265–273.
- Peuskens J, Risperidone Study Group (1995). Risperidone in the treatment of patients with chronic schizophrenia: a multinational, multi-centre, double-blind, parallel-group study versus haloperidol. *Br J Psychiatry* 166: 712–726.
- Puech A, Fleurot O, Rein W (1998). Amisulpride, an atypical antipsychotic, in the treatment of acute episodes of schizophrenia: a dose-ranging study vs. haloperidol. *Acta Psychiatr Scand* 98: 65–72.
- Sèchter D, Peuskens J, Fleurot O, Rein W, Lecrubier Y (2002). Amisulpride vs risperidone in chronic schizophrenia: results of a 6-month double-blind study. *Neuropsychopharmacology* 27: 1071–1081.
- Small JG, Hirsch SR, Arvanitis LA, Miller BG, Link CGG, Seroquel Study Group (1997). Quetiapine in patients with schizophrenia. A high- and low-dose comparison with placebo. *Arch Gen Psychiatry* 54: 549–557.
- Thompson PA, Buckley PF, Meltzer HY (1994). The brief psychiatric rating scale: effect of scaling system on clinical response assessment. *J Clin Psychopharmacol* 14: 344–346.
- Tollefson GD, Birkett MA, Kiesler GM, Wood AJ, Lilly Resistant Schizophrenia Study Group (2001). Double-blind comparison of olanzapine versus clozapine in schizophrenic patients clinically eligible for treatment with clozapine. *Biol Psychiatry* 49: 52–63.
- Wang SL, Weng YZ, Huang SZ, Han XL, Li KJ (2001). A comparison of clozapine to chlorpromazine affecting serum glutamate levels and their relationship with clinical efficacy in schizophrenia. *Chin J Behav Med Sci* 10: 565–567.
- Wetzel H, Grunder G, Hillert A, Philipp M, Gattaz WF, Sauer H et al (1998). Amisulpride versus flupentixol in schizophrenia with predominantly positive symptomatology—a double-blind controlled study comparing a selective D-2-like antagonist to a mixed D-1-/D-2-like antagonist. *Psychopharmacology* 137: 223–232.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>)