



Published in final edited form as:

J Nerv Ment Dis. 2014 January ; 202(1): 18–24. doi:10.1097/NMD.0000000000000065.

Noninferiority of Perphenazine vs. Three Second-Generation Antipsychotics in Chronic Schizophrenia

Robert Rosenheck, MD^{*,†,‡} and Haiqun Lin, PhD[‡]

^{*}VA New England Mental Illness Research and Education Center, West Haven, CT

[†]Department of Psychiatry, Yale University School of Medicine, New Haven, CT

[‡]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT

Abstract

Noninferiority analysis is a statistical method of growing importance in comparative effectiveness research that has rarely been used in psychopharmacology. This method is used here to evaluate whether first-generation antipsychotics are clinically not inferior to second-generation antipsychotics (SGAs) using data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). A conservative noninferiority margin (NIM) on the Positive and Negative Syndrome Scale (PANSS) was derived from the smallest published value for the minimal clinically important difference, further reduced by 25%. This NIM was used to assess whether perphenazine is noninferior to olanzapine, risperidone, and quetiapine on the basis of the 95% confidence intervals of differences in mean PANSS outcomes ($N = 1049$). Perphenazine was noninferior to all three SGAs during 18 months of intention-to-treat analysis and in several subanalyses. Noninferiority can be evaluated from studies designed as superiority trials. Power was available in the CATIE to conduct noninferiority analysis.

Keywords

Antipsychotics; noninferiority analysis; schizophrenia

Comparative effectiveness research on the relative benefits of approved medications such as second-generation antipsychotics (SGAs; *e.g.*, olanzapine, risperidone, and quetiapine) and first-generation antipsychotics (FGAs) such as perphenazine and haloperidol has been characterized by apparently inconsistent results, with some studies showing some SGAs to be superior to FGAs (Leucht et al., 2009), whereas others fail to find such superiority (Rosenheck and Sernyak, 2009). This has led to conflicting and sometimes controversial interpretations (Kraemer et al., 2009; Tyrer and Kendall, 2009). As SGAs become generic and FGA-SGA cost differences recede, the issue of relative clinical effectiveness will

Send reprint requests to Robert Rosenheck, MD, 151D VA New England Mental Illness, Research, Education and Clinical Center/ 151D VA Connecticut Health Care System, 950 Campbell Ave, Building 35, West Haven, CT. robert.rosenheck@yale.edu.

Both authors were responsible for the conception and design, analysis, and interpretation of data.

DISCLOSURES

The authors declare no conflict of interest.

become more central to clinical decision making. Although numerous short-term trials have shown superior efficacy for SGAs as compared with FGAs (Leucht et al., 2009), more recent longer-term comparative effectiveness trials such as the US Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE; Lieberman et al., 2005), the UK Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS; Jones et al., 2006), and a Veterans Affairs (VA) cooperative study (Rosenheck et al., 2003) failed to find evidence of superior efficacy for SGAs. As noted by Leucht et al. (2009), the CATIE and the CUtLASS focused on real-world effectiveness, and an additional strength of the CATIE was the use of perphenazine, an intermediate-potency agent, rather than haloperidol, a high-potency agent, as the comparator, as in most other trials.

Commentaries on the CATIE have underlined the principle that failure to prove superiority cannot be taken as evidence of noninferiority or equivalence, that is, failing to reject the null hypothesis that a particular SGA is not different in benefits from a particular FGA does not, in itself, support the conclusion that the treatments are clinically equivalent (Freedman et al., 2006; Kraemer et al., 2009; Kraemer, 2011; Leon, 2011). Noninferiority analysis, a method rarely applied in comparative psychopharmacology trials, is needed to address this issue (Gerligner and Schmelter, 2011). Virtually all published trials of SGAs are based on the statistical premise that failure to reject the null hypothesis, that is, finding that the 95% confidence interval (CI) of the difference between treatments includes zero, rules out the possibility that one treatment is superior to the other. However, such results cannot be taken as evidence that FGAs are clinically noninferior than SGAs without further analysis. Noninferiority analysis, in contrast, requires that the 95% CI of the difference in means be smaller than an independently determined clinically meaningful difference, referred to as the noninferiority margin (NIM; Leon, 2011).

Failure to reject the null hypothesis can indicate either that one treatment is not inferior to another within an NIM or that there was insufficient statistical power or sample size to detect a difference between treatments. Thus, although none of four SGAs was superior to the FGA perphenazine on the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987) in the CATIE, the question of whether perphenazine or any other FGA is noninferior or clinically equivalent to SGA drugs requires a specific noninferiority analysis (Kraemer et al., 2009; Kraemer, 2011; Leon, 2011). The goal of this study was to outline the principles of noninferiority analysis and apply them to data from the CATIE.

The CATIE was originally powered for the primary outcome of time from randomization to all-cause medication discontinuation and assumed that a 15% difference in all-cause discontinuation at 18 months would be clinically meaningful (expected to be 30% for SGAs vs. 45% for perphenazine; Lieberman et al., 2005). Although discontinuation has been shown to be associated with poorer outcomes (Davis et al., 2011), no plan was included in the original design to assess noninferiority as defined above. In this article, we focus on comparing psychotic symptom outcomes between treatments on the PANSS. The minimal clinically important difference (MCID) has been defined as “the smallest difference in a score in the domain of interest which patients [or providers] perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a meaningful change in the patient’s management” (Jaeschke et al., 1989). We first

independently derive an NIM for the PANSS using published assessments of MCID on the PANSS (Cramer et al., 2001; Hermes et al., 2012; Leucht et al., 2005; Levine et al., 2008; Rabinowitz et al., 2006; Schennach-Wolff et al., 2010; Thwin et al., in press), as described in detail below, and use this NIM to evaluate whether perphenazine is noninferior to the three SGAs originally included in the CATIE. The NIM is often (although not always) smaller than the MCID (Leon, 2011). The use of MCID in noninferiority analysis has been demonstrated for other medical conditions (Jaeschke et al., 1989).

We acknowledge that noninferiority assessment represents a secondary analysis but note that, as in a cancer trial, “An unplanned non-inferiority conclusion sometimes occurs when the results of a superiority trial were negative but the data suggested that a noninferiority argument can be applied...” (Zee, 2006). Because noninferiority analyses have not been used in comparisons of FGAs and SGAs, and the approach has rarely been used in psychiatric research, we think that such analyses of CATIE data can be informative. We use the smallest MCID in the literature for the PANSS in schizophrenia because it is the most conservative basis for calculating the NIM. A smaller NIM requires a larger sample size and/or a smaller difference between treatments to conclude noninferiority. Use of the smallest published MCID is considered conservative because it makes the strongest effort to avoid considering a medication noninferior when it is not, although reciprocally, it risks failing to demonstrate noninferiority where it could otherwise be claimed. Particularly in the case of FGAs, which many clinicians regard as inferior to SGAs, we think that such a conservative approach is indicated. Therefore, if we demonstrate noninferiority using the smallest MCID as the basis for the NIM, the conclusion of noninferiority clearly follows for larger estimates of the MCID, as described below.

METHODS

Study Setting and Design

The CATIE schizophrenia trial was conducted, between January 2001 and December 2004, at 57 US sites and included an algorithmically determined series of treatment phases (Lieberman et al., 2005). Patients 18 to 65 years of age diagnosed with schizophrenia were initially randomized to the FGA, perphenazine, or to one of three SGAs, olanzapine, quetiapine, or risperidone, under double-blind conditions. Patients with tardive dyskinesia were prohibited from assignment to randomizations that included perphenazine. This group of 1049 patients assigned to perphenazine ($n = 256$), olanzapine ($n = 263$), quetiapine ($n = 261$), or risperidone ($n = 269$) is the focus of the analyses presented here. Patients who discontinued their first (phase 1) treatment were invited to receive other SGAs, including clozapine if they so desired, with either random assignment to specific agents or open treatment. Ziprasidone, a fourth SGA, entered the trial after 40% of the sample had been recruited, and thus offers a smaller sample size, and is not examined here.

Methodological details including the CONSORT flow diagram have been published previously (Lieberman et al., 2005). Identical capsules contained olanzapine (7.5 mg), quetiapine (200 mg), risperidone (1.5 mg), perphenazine (8 mg), or ziprasidone (40 mg). Medications were flexibly dosed with one to four capsules daily, as judged by the study

physician. This study was approved by an institutional review board at each site and registered at [ClinicalTrials.gov](https://clinicaltrials.gov) as NCT00014001. Both authors had full access to the data.

Outcomes

Symptoms of schizophrenia were measured by the PANSS total score (range, 30–210), with higher scores indicating more symptoms (Kay et al., 1987). PANSS assessments were conducted at baseline and 1, 3, 6, 9, 12, and 18 months after randomization.

Determination of the MCID and NIM

The concept of the MCID is an important basis for evaluating the clinical importance of differences between treatments and stands in contrast to the usual statistical significance testing based only on whether the 95% CI of the difference between treatments excludes zero (illustrated by CIs 1–6 in Figure 1, all of which exclude 0). Clinical superiority is typically based on the MCID, whereas noninferiority is more conservatively based on the NIM, which, it has been argued, should be smaller than the MCID, although there is no analytic procedure for determining how much smaller the NIM should be, making such reductions somewhat arbitrary (Leon, 2011). We feel that this reduction of the NIM to a proportion less than the MCID is justified because more clinical harm would come from erring on the side of considering an inferior treatment to be noninferior, than the reverse, because it would increase the risk of patients being exposed to an inferior treatment. In this article, we use a conservative, and admittedly somewhat arbitrary, NIM based on the smallest MCID in the published literature for the PANSS further reduced by 25%.

Several studies have used what are called “anchor-based” methods to determine the MCID of the PANSS in patients with schizophrenia. These methods use a measure with face-value clinical meaning such as the Clinical Global Impressions (CGI) Severity Scale for cross-sectional assessment or the CGI Improvement Scale (CGI-I) for improvement over time to anchor a statistical evaluation of clinically meaningful scores on the PANSS. Using the CGI, a clinician or patient comes to a global rating of overall severity of illness or change in severity of the illness since baseline on a 7-point scale. The intervals on this scale (especially the interval between no change and either minimally improved or minimally worsened) are taken as representing the smallest clinically differentiable or meaningful differences and can be used to determine the number of PANSS points that separate these levels of severity.

Assessments of the MCID for percent changes in the PANSS in relation to the CGI-I have estimated that a change in standard PANSS scores of between 17% and 24% corresponds to the MCID (Cramer et al., 2001; Hermes et al., 2012; Leucht et al., 2005; Levine et al., 2008; Rabinowitz et al., 2006; Schennach-Wolff et al., 2010; Thwin et al., in press). These percentages compute to 12 to 18 PANSS scale points in the case of the CATIE. Published estimates of the MCID using PANSS scale points directly cluster around a 15-point estimate of the MCID (Hermes et al., 2012; Leucht et al., 2005; Schennach-Wolff et al., 2010), with an additional overall estimate of 18.6 PANSS points from a study of hospitalized VA patients at 6 and 12 months (Thwin et al., in press). However, the VA study also presents more specific data on PANSS changes between CGI ratings of no change and minimally

improved status (−10.7 PANSS points) and between CGI ratings of no change and minimally worse status (+8.4 points), which, at face value, are more likely to represent minimal clinical differences, although there is thus a broad range from 8.4 to 18.6 PANSS points for the estimated MCID of the PANSS, with more evidence for the higher (>12) range of the scale. We use the smallest, and therefore most conservative, published MCID estimate of 8.4 PANSS. Because it is recommended that a figure lower than the MCID be used as the NIM (Leon, 2011), we reduced this smallest MCID by 25%, from 8.4 to 6.3 PANSS points.

Assessment of Noninferiority

Because lower scores reflect better outcomes on the PANSS, we measure the difference in effect as the mean score for perphenazine at follow-up (least square means adjusted for baseline value of the dependent variable), the control condition, minus the mean score for each SGA (the “experimental” conditions). Least square means are used to present mean outcome values that are adjusted for any differences (even nonsignificant differences) in baseline values.

If the 95% CI of the difference of perphenazine and an SGA lies within the interval between the negative NIM and the positive NIM, one can be statistically confident that the difference between the treatments is not clinically meaningful, that is, the treatments are clinically equivalent or noninferior to each other. The noninferiority of perphenazine to an SGA and the noninferiority of an SGA to perphenazine are illustrated heuristically by CIs 5 to 7 in Figure 1.

If the upper bound of the 95% CI of the difference between perphenazine and an SGA is lower than the positive NIM (CIs 1, 3, and 5 to 8 in Figure 1), we could still conclude that perphenazine is noninferior to the SGA because we can be 95% sure that the SGA is not superior to perphenazine with a magnitude greater than the NIM. Reciprocally for CIs 2, 4 to 7, and 9 in Figure 1, we could conclude that the SGA is noninferior to perphenazine because we can be 95% sure that perphenazine is not superior to the SGA with a magnitude less than the NIM.

To conclude that a treatment is clinically superior, the 95% CI of the difference between treatments must to lie entirely outside the interval of the MCID range (*e.g.*, CI 3 favoring perphenazine and CI 4 favoring the SGA in Figure 1). CIs 1, 2, 5, and 6 in Figure 1, in contrast, represent statistically significant superiority (because these do not include 0) but not clinically meaningful superiority. CI 10 in Figure 1 fails to show either clinical superiority for either treatment or noninferiority.

A larger estimate of the MCID as the basis for the NIM is more likely to encompass the 95% CIs of the difference between treatments than a smaller one and may thus allow conclusions of noninferiority, even in studies with modest sample sizes. If the NIM in Figure 1 was slightly larger, CIs 9 and 10 might represent noninferiority of perphenazine.

Statistical Analysis

Mixed Model Analysis—For consistency and comparability, statistical methods used in the analysis of continuous measures in this study were patterned on those used in the original publications from the CATIE trial (Lieberman et al., 2005; Rosenheck et al., 2006). The mean PANSS scores over all available observations during 18 months were compared between paired treatments; with a mixed model including terms representing treatment assignment as a class variable, we additionally adjusted for time (treated as a class variable for months 1–18), treatment by time interaction, site, a history of recent clinical exacerbation, the baseline value of the PANSS for each patient, and baseline-by-time interactions. The baseline-by-time term adjusts for baseline differences in characteristics of patients who dropped out early and thus are less well represented at later time points. A random subject effect is used to account for individual patient effects. A firstorder autoregressive covariance structure was used for the analyses of all time points and for the secondary analyses of months 1 to 6, in which there are three or more designed times for measuring the PANSS. An independent covariance structure was used for an analysis of time points 9 to 12 and 15 to 18 months, in which there were only two time points.

We present adjusted mean PANSS score for each treatment along with the difference between treatments on the PANSS, represented as perphenazine minus each SGA, and the overall 95% CIs of these differences, adjusted for three paired comparisons using the conservative Bonferroni's correction. These data are presented for both the intention-to-treat (ITT) sample (including all observation after randomization) and the "phase 1-only" samples (excluding observations after a change from the originally randomized medication to another drug) using data across the entire study. In addition, we present ITT and phase 1-only comparisons of data collected from 1 to 6 months, 9 to 12 months, and 15 to 18 months.

RESULTS

There were no significant differences between the groups on the PANSS total score or other sociodemographic or clinical measures at the time of randomization (Table 1).

On both the overall 18-month ITT analysis ($n = 5852$ observations from 1049 patients) and the phase 1-only analysis ($n = 4453$ observations from 1047 patients), perphenazine was noninferior to each SGA, with all 95% CIs less than 4.0 PANSS points, less than two thirds of the estimate for the NIM (Table 2). In the ITT analysis, olanzapine was statistically but not clinically superior to perphenazine, whereas perphenazine was statistically but not clinically superior to risperidone.

In both ITT and phase 1-only analyses of comparisons at 1 to 6 months, 9 to 12 months, and 15 to 18 months, perphenazine was also demonstrated to be noninferior to each SGA. In both ITT and phase 1-only analyses of comparisons at 1 to 6 months, as in the overall ITT analysis, olanzapine was statistically but not clinically superior to perphenazine, whereas perphenazine was statistically but not clinically superior to risperidone (Table 2). In the phase 1-only analysis at 9 to 12 months, perphenazine was also statistically but not clinically superior to risperidone.

DISCUSSION

This study used data from the CATIE schizophrenia trial and the smallest of several published estimates for the MCID for the PANSS total score to identify an NIM for the PANSS. It then applied the principles of noninferiority analysis to the comparison of the effectiveness of the FGA perphenazine with that of three widely used SGAs. These analyses have both methodological and substantive importance. Methodologically, these demonstrate that data from an effectiveness trial of schizophrenia treatments can be evaluated from a noninferiority framework and that, at least in this case, power was sufficient to demonstrate noninferiority. Substantively, these analyses suggest that data from the CATIE show perphenazine to have been noninferior to olanzapine, quetiapine, and risperidone.

We are not the first to use the MCID to derive the NIM for noninferiority analysis. In 2011, Gerlinger and Schmelter derived the NIM in several medical conditions from MCIDs estimated using the anchor-based method (virtually the same approach as that used here). In fact, they set the NIM to be equal to the MCID, a less conservative strategy than the further 25% reduction of the MCID used here as recommended by Leon (2011). Although we also adopted the more conservative strategy for the evaluation of FGAs and SGAs, we would note that making the NIM smaller is not always more appropriate, for example, when a drug with less severe side effects is compared with a drug that is regarded as more effective in clinical trials but has more side effects.

Further Applications

The approach to clinical importance presented here can also be applied to published data from the two landmark superiority trials that found statistically significant benefits for the SGAs risperidone and olanzapine as compared with haloperidol, a higher-potency FGA than perphenazine.

The first major publication comparing risperidone at various doses with haloperidol (Marder and Meibach, 1994) found risperidone to be statistically significantly better than haloperidol at both 6-mg and 16-mg doses although not at the intermediate dose of 10 mg. Data presented in the original publication allow calculation of 95% CIs for the mean differences between treatments on the PANSS. At 8 weeks, the mean difference between treatments for haloperidol minus the 6-mg dose of risperidone was 12.0 (95% CI, 5.29–18.71), whereas at 16 mg, it was 9.4 (95% CI, 3.7–15.17), both of which are representative of CI 2 in Figure 1. Although statistically significant, neither of these estimates allows a conclusion of clinical superiority for risperidone, in part because the small sample sizes ($n = 62$ – 63 per treatment group) yield wide CIs—considerably wider than were found in our analysis of CATIE data. These data do allow the conclusion that risperidone is noninferior to haloperidol, although it is not shown to be clinically superior. It does not show that haloperidol is noninferior to risperidone.

In the comparison of olanzapine with haloperidol in the much larger ($N = 1996$) International Collaborative Trial (Tollefson et al., 1997), the mean difference (FGA minus SGA) on the PANSS total score was 4.3 (95% CI, 2.4–6.2). This is a much smaller 95% CI than in the risperidone trial because of the larger sample size. These data also fail to meet

our criterion for the clinical superiority of olanzapine. In fact, this 95% CI does meet our criteria for the mutual noninferiority (*i.e.*, clinical equivalence) of haloperidol and olanzapine (represented by CI 6 in Figure 1w although the difference was statistically significant in favor of olanzapine, albeit at $p = 0.05$ with a small effect size of 0.2 (Tollefson et al., 1997). Thus, although the CATIE and these earlier studies seem to have conflicting results using standards of statistical significance applied, these yield more consistent conclusions even when the most conservative estimate of the MCID is used as the basis for evaluating clinical meaningfulness, that is, conservative in the sense of avoiding considering medication noninferior, when these might actually be inferior.

The major limitation of this study is that it was a secondary descriptive analysis designed after the data were collected and the superiority analysis was completed (Zee, 2006). However, our NIM is derived independently of the published superiority analysis of the CATIE and is consistent with the original study results that showed no superiority of any SGA to perphenazine on the primary outcome of time to all-cause discontinuation after adjustment for multiple comparisons (Lieberman et al., 2005).

A potential additional limitation is that none of the MCID studies cited demonstrated that the difference between intervals on the CGI represents a definitive anchor for estimating the minimal discernable difference in clinical status. In response to this uncertainty, we used the lowest estimate of 8.4 PANSS points for the CGI-I difference between no change and minimally improved and further reduced this MCID value by 25% for an NIM of 6.3, substantially lower than the predominant range of published estimates for the PANSS MCID of 12 to 18 PANSS points and thus a quite conservative figure.

A limitation of using the PANSS for noninferiority analysis is that it assesses only symptoms and not side effects, and a comprehensive assessment of noninferiority would need to consider adverse effects and symptoms. As far as the CATIE study is concerned, numerous substudies of side effects, social and neurocognitive functioning, employment, and violent behavior all found no statistically significant difference between perphenazine and any SGA with the exception of greater weight gain and metabolic risk with olanzapine (Stroup and Lieberman, 2010).

Finally, we acknowledge that noninferiority analysis is only one approach to evaluating clinically meaningful differences in clinical trials. A recent article used a published standard for identifying “remission” in schizophrenia to determine the differences in the proportions of patients achieving remission on different treatments in schizophrenia and in the CATIE trial in particular, finding no superiority for any SGA over perphenazine during 3- or 6-month remission periods, after adjustment for multiple comparisons (Levine et al., 2011).

CONCLUSIONS

This study shows that noninferiority can be evaluated using data from studies originally designed as superiority trials when sample sizes are sufficiently large. Furthermore, ample power was available in the CATIE trial to show that perphenazine was not inferior to olanzapine, quetiapine, or risperidone.

REFERENCES

- Cramer J, Rosenheck R, Xu W, Henderson W, Thomas J, Charney D. Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. Improvement in quality of life and symptomatology in schizophrenia. *Schizophr Bull.* 2001; 27:227–234. [PubMed: 11354590]
- Davis S, Stroup S, Koch E, Davis E, Rosenheck RA, Lieberman J. Time to all-cause treatment discontinuation as the primary outcome in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Schizophrenia Study. *Stat Biopharm Res.* 2011; 3:253–265.
- Freedman R, Carpenter WT, Davis JM, Goldman HH, Tamminga CA, Thomas M. The costs of drugs for schizophrenia. *Am J Psychiatry.* 2006; 163:2029–2031. [PubMed: 17151146]
- Gerlinger C, Schmelter T. Determining the non-inferiority margin for patient reported outcomes. *Pharm Stat.* 2011; 10:410–413. [PubMed: 21932298]
- Hermes E, Sokoloff D, Rosenheck RA. Minimum clinically important difference in the Positive and Negative Syndrome Scale using data from the CATIE schizophrenia trial. *J Clin Psychiatry.* 2012; 73:526–532. [PubMed: 22579152]
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989; 10:407–415. [PubMed: 2691207]
- Jones PB, Barnes TR, Davies L, Dunn G, Lloyd H, Hayhurst KP, Murray RM, Markwick A, Lewis SW. Randomized controlled trial of effect on quality of life of second-generation versus first-generation antipsychotic drugs in schizophrenia—CUtLASS 1. *Arch Gen Psychiatry.* 2006; 63:1079–1087. [PubMed: 17015810]
- Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987; 13:261–276. [PubMed: 3616518]
- Kraemer H, Glick ID, Klein DF. Clinical trials design lessons from the CATIE Study. *Am J Psychiatry.* 2009; 166:1222–1228. [PubMed: 19797435]
- Kraemer HC. Another point of view: Superiority, noninferiority, and the role of active comparators. *J Clin Psychiatry.* 2011; 72:1350–1352. [PubMed: 22075101]
- Leon A. Comparative effectiveness clinical trials in psychiatry: Superiority, non-inferiority and the role of active comparators. *J Clin Psychiatry.* 2011; 72:331–340. [PubMed: 21450153]
- Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: A meta-analysis. *Lancet.* 2009; 373:31–41. [PubMed: 19058842]
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophr Res.* 2005; 15:231–238. [PubMed: 15982856]
- Levine SZ, Rabinowitz J, Ascher-Svanum H, Faries DE, Lawson AH. Extent of attaining and maintaining symptom remission by antipsychotic medication in the treatment of chronic schizophrenia: Evidence from the CATIE study. *Schizophr Res.* 2011; 133:42–46. [PubMed: 22000938]
- Levine SZ, Rabinowitz J, Engel R, Etschel E, Leucht S. Extrapolation between measures of symptom severity and change: An examination of the PANSS and CGI. *Schizophr Res.* 2008; 98:318–322. [PubMed: 17949948]
- Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, Keefe RS, Davis SM, Davis CE, Lebowitz BD, Severe J, Hsiao JK. Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med.* 2005; 353:1209–1223. [PubMed: 16172203]
- Marder SR, Meibach RC. Risperidone in the treatment of schizophrenia. *Am J Psychiatry.* 1994; 151:825–834. [PubMed: 7514366]
- Rabinowitz J, Mehnert A, Eerdeken M. To what extent do the PANSS and CGI-S overlap? *J Clin Psychopharmacol.* 2006; 26:303–307. [PubMed: 16702895]
- Rosenheck RA, Leslie D, Sindelar J, Miller EA, Lin H, Stroup S, McEvoy J, Davis S, Keefe RSE, Swartz M, Perkins D, Hsiao J, Lieberman JA. Costeffectiveness of second generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia. *Am J Psychiatry.* 2006; 163:2080–2089. Supporting data: Data supplement: http://ajp.psychiatryonline.org/data/Journals/AJP/3787/AJP_163_12_2080_01.pdf. [PubMed: 17151158]

- Rosenheck RA, Perlick D, Bingham S, Liu-Mares W, Collins J, Warren S, Leslie D. for the Department of Veterans Affairs Cooperative Study Group on the Cost- Effectiveness of Olanzapine. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia. *JAMA*. 2003; 290:2693–2702. [PubMed: 14645311]
- Rosenheck RA, Sernyak MJ. Developing a policy for second-generation antipsychotic drugs. *Health Affairs*. 2009; 28:w782–w793. Published online July 21, 2009. Retrieved from <http://content.healthaffairs.org/cgi/reprint/hlthaff.28.5.w782v1>. [PubMed: 19622538]
- Schennach-Wolff R, Obermeier M, Seemüller F, Jäger M, Schmauss M, Laux G, Pfeiffer H, Naber D, Schmidt LG, Gaebel W, Klosterkötter J, Heuser I, Maier W, Lemke MR, Rütter E, Klingberg S, Gastpar M, Engel RR, Möller HJ, Riedel M. Does clinical judgment of baseline severity and changes in psychopathology depend on the patient population? Results of a CGI and PANSS linking analysis in a naturalistic study. *J Clin Psychopharmacol*. 2010; 30:726–731. [PubMed: 21105273]
- Stroup, TS.; Lieberman, JA., editors. Antipsychotic trials in schizophrenia: The CATIE Project. Cambridge, England: Cambridge University Press; 2010.
- Thwin SS, Hermes E, Lew R, Barnett P, Liang M, Valley D, Rosenheck RA. Assessment of the minimum clinically important difference in quality of life in schizophrenia measured by the Quality of Well-Being Scale and disease-specific measures. *Psychiatry Res*. (in press).
- Tollefson GD, Beasley CM Jr, Tran PV, Street JS, Krueger JA, Tamura RN, Graffeo KA, Thieme ME. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: Results of an international collaborative trial. *Am J Psychiatry*. 1997; 154:457–465. [PubMed: 9090331]
- Tyrer P, Kendall T. The spurious advance of antipsychotic drug therapy. *Lancet*. 2009; 373:4–5. [PubMed: 19058841]
- Zee BC. Planned equivalence or noninferiority trials versus unplanned noninferiority claims: Are they equal? *J Clin Oncol*. 2006; 24:1026–1028. [PubMed: 16505420]

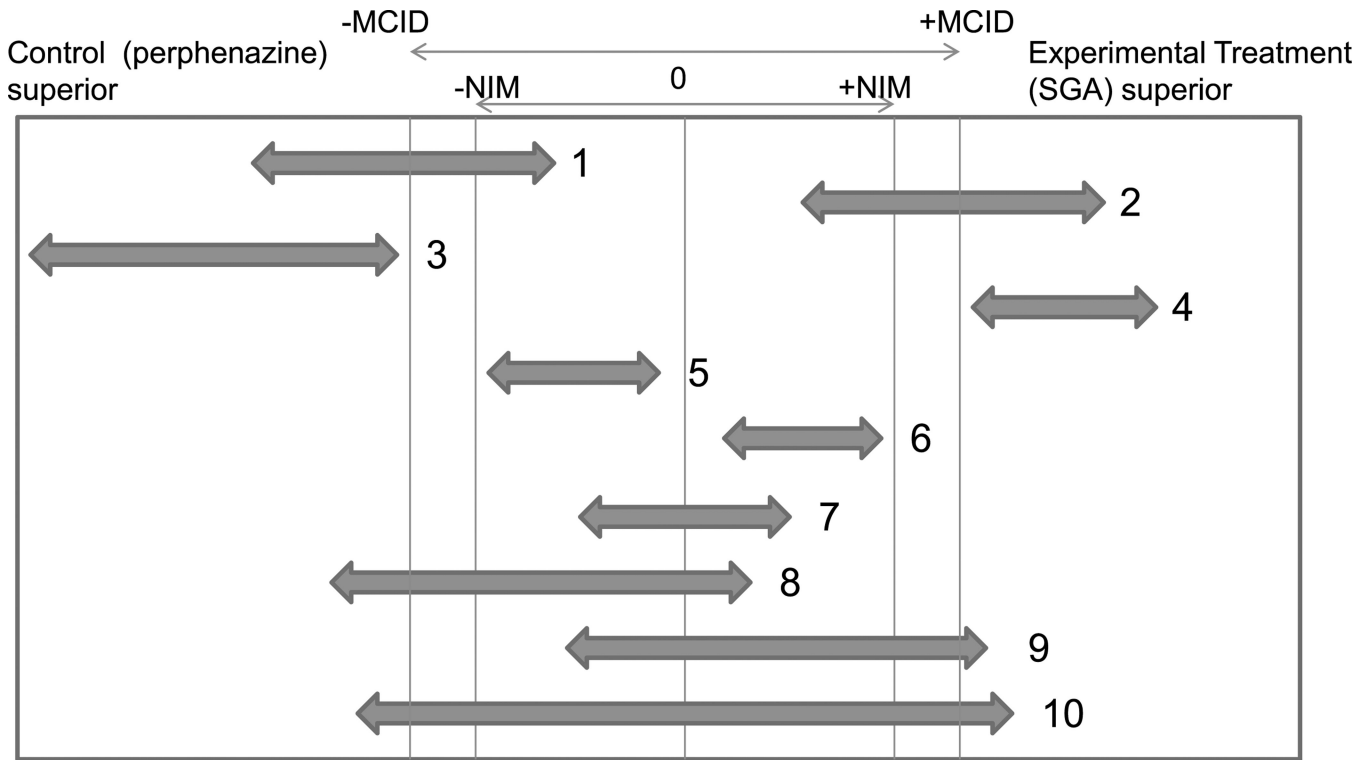


FIGURE 1. Interpretation of clinically and statistically meaningful differences between treatments. The two-way arrows represent two-sided 95% CIs of differences between treatment groups defined as control minus experimental. The smallest MCID reduced by 25% to represent the NIM was used to evaluate the clinical relevance of treatment differences in contrast to the statistical significance test, which determines whether zero lies outside the 95% CI (CIs 1–6). If the 95% CI falls between the –NIM and the + NIM, one can be statistically confident that the treatments are mutually noninferior to each other (CIs 5, 6, and 7). Furthermore, when a lower measure of the effect is considered a better outcome, as with most symptom measures, when the upper confidence bound of the difference between treatments is lower than the +NIM, we can still conclude that the control is noninferior to the experimental treatment (CIs 1, 3, and 5–8) but not vice versa. Clinically meaningful superiority requires that the 95% CI lies entirely outside the interval of the range –MCID to + MCID (CIs 3 and 4). CIs 1, 2, 5, and 6 thus represent statistically significant superiority but not clinically meaningful superiority.

TABLE 1
Patient Characteristics at the Time of Random Assignment by Initial Treatment for the Randomization That Excluded Patients With TD or Those Assigned to Ziprasidone

	Total Sample		Olanzapine		Perphenazine		Quetiapine		Risperidone		χ^2/F	df	p
	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD			
	n = 1049		n = 263		n = 256		n = 261		n = 269				
Age	39.3	10.9%	39.36	10.56	39.97	11.06	39.15	10.88	38.78	11.05	0.55	3	0.65
Male	777	74.0%	190	72.2%	196	76.6%	196	73.5%	199	74.0%	1.3	3	0.72
Race/ethnicity													
White	631	60.1%	153	58.2%	151	59.0%	167	64.0%	160	59.5%	2.22	3	0.52
Black	368	35.1%	96	36.5%	90	35.1%	84	32.1%	98	36.4%	1.4	3	0.70
Other	50	4.8%	14	5.3%	15	5.9%	10	3.8%	11	4.1%	1.62	3	0.65
Hispanic	129	12.3%	37	14.1%	24	9.3%	39	14.9%	29	10.8%	5.05	3	0.17
Marital status											16.7	12	0.16
Married	131	12.5%	30	11.4%	43	16.8%	27	10.3%	31	11.5%			
Separated/divorced	219	20.8%	61	23.2%	50	19.4%	55	20.9%	53	19.8%			
Never married	636	60.6%	159	60.4%	146	57.0%	167	64.0%	164	61.0%			
Widowed													
PANSS total	75.5	17.5	75.7	18.2	74.2	18.0	74.8	17.0	77.2	16.5	1.48	3	0.21
Positive	18.4	5.6	18.4	5.5	17.9	5.9	18.3	5.4	19.0	5.6	1.8	3	0.15
Negative	20.2	6.5	20.3	6.7	20.3	6.3	19.8	6.5	20.4	6.4	0.43	3	0.72
General	36.9	9.3	37.0	9.8	36.0	9.5	36.7	9.2	37.8	8.6	1.68	3	0.17
Depression (Cagary Scale)	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.6	0.84	3	0.47
Current comorbidity													
Alcohol abuse/dependence	86	8.2%	20	7.6%	27	10.5%	21	8.0%	18	6.7%	2.83	3	0.41
Drug abuse/dependence	126	11.9%	30	11.3%	39	15.2%	23	8.8%	34	12.6%	5.31	3	0.15
Major depression	114	10.8%	25	9.4%	29	11.3%	23	8.8%	37	13.7%	4.08	3	0.25
Obsessive-compulsive disorder	57	5.4%	8	3.0%	11	4.3%	20	7.7%	18	6.7%	6.9	3	0.07
Side effects											3		
EPS mean (Simpson Agua)	0.18	0.29	0.16	0.27	0.19	0.32	0.16	0.25	0.20	0.29	1.35	3	0.26

	Total Sample		Olanzapine		Perphenazine		Quetiapine		Risperidone		χ^2/F	df	p
	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD	Mean/n	%/SD			
	n = 1049		n = 263		n = 256		n = 261		n = 269				
Barnes Akathisia Scale	0.47	0.84	0.58	0.97	0.43	0.79	0.46	0.76	0.47	0.81	2.17	3	0.09
AIMS Severity Score (TD)	0.12	0.27	0.15	0.31	0.12	0.27	0.11	0.24	0.12	0.23	1.13	3	0.34
TD (>1 on AIMS Severity)	7	67.0%	1	0.4%	3	1.2%	2	0.8%	1	0.4%	1.7	3	0.63
Quality of Life Scale	2.7	1.1	2.7	1.1	2.7	1.1	2.7	1.1	2.6	1.0	0.73	3	0.53
Social interaction	2.6	1.3	2.5	1.2	2.6	1.4	2.6	1.4	2.5	1.3	1.21	3	0.30
Instrumental activity	2.0	1.6	2.0	1.7	2.0	1.6	2.0	1.7	1.9	1.5	0.19	3	0.91
Intrapsychic activity	3.0	1.2	3.1	1.2	3.0	1.2	3.1	1.2	2.9	1.0	1.35	3	0.26
Lehman Quality of Life Interview	4.3	1.4	4.4	1.4	4.4	1.3	4.4	1.4	4.2	1.5	0.83	3	0.47
Quality-adjusted life-years	0.686	0.126	0.682	0.132	0.689	0.120	0.695	0.123	0.676	0.127	1.11	3	0.34
Patient weighted health index	-0.94	3.64	-1.12	3.98	-0.75	3.67	-0.59	3.43	-1.29	3.41	1.86	3	0.13
Body mass index	29.80	7.09	29.24	6.86	29.63	6.93	30.22	7.05	30.09	7.48	1.04	3	0.37
Health costs (previous month)													
All medication	\$422	\$325	\$419	\$344	\$420	\$314	\$418	\$331	\$433	\$313	0.14	3	0.94
Inpatient residential/nursing home	\$1512	\$3715	\$1828	\$3988	\$1127	\$2530	\$1442	\$3642	\$1636	\$4381	1.68	3	0.17
Outpatient	\$365	\$935	\$379	\$864	\$392	\$1173	\$410	\$1066	\$281	\$513	1.02	3	0.38
Total	\$2299	\$3831	\$2628	\$4078	\$1940	\$2811	\$2271	\$3813	\$2352	\$4398	1.42	3	0.23

AIMS indicates abnormal involuntary movement scale; EPS, extrapyramidal syndrome; TD, tardive dyskinesia.

TABLE 2

Comparison of PANSS Total Scores Between Perphenazine and Three SGAs With 95% CIs of Differences in LS Means^a

Treatment	LS Mean of PANSS Total Score	Difference from Perphenazine	SD of Difference in LS Mean	95% CI of Group Differences: Lower Limit	95% CI of Group Differences: Upper Limit	p ^b
ITT analysis						
All data during 18 mos (n = 5852 observations from 1047 patients)						
Perphenazine	67.72					
Olanzapine	65.93	1.79	0.74	-0.04	3.54	
Quetiapine	68.02	-0.30	0.76	-2.08	1.49	
Risperidone	69.64	-1.92	0.75	-3.70	-0.14	*
Months 1, 3, and 6 (n = 2492 observations from 978 patients)						
Perphenazine	69.25					
Olanzapine	67.14	2.11	0.90	-0.03	4.25	
Quetiapine	69.78	-0.53	0.92	-2.71	1.64	
Risperidone	71.80	-2.55	0.92	4.72	-0.39	*
Months 9 and 12 (n = 1273 observations from 677 patients)						
Perphenazine	64.84					
Olanzapine	63.97	0.87	1.35	-2.32	4.07	
Quetiapine	65.89	-1.05	1.42	-4.40	2.30	
Risperidone	67.10	-2.26	1.41	-5.57	1.06	
Months 15 and 18 (n = 1054 observations from 578 patients)						
Perphenazine	63.52					
Olanzapine	61.12	2.40	1.52	-1.19	5.98	
Quetiapine	63.25	0.27	1.61	-3.53	4.08	
Risperidone	63.62	-0.10	1.62	-3.92	3.71	
Phase 1 only						
All data during 18 mos (n = 4453 observations from 1047 patients)						
Perphenazine	66.34					
Olanzapine	65.25	1.09	0.75	-0.67	2.85	
Quetiapine	67.89	-1.55	0.79	-3.41	0.31	
Risperidone	68.52	-2.18	0.77	-4.00	-0.36	
Months 1, 3, and 6 (n = 2122 observations from 973 patients)						

Treatment	LS Mean of PANSS Total Score	Difference from Perphenazine	SD of Difference in LS Mean	95% CI of Group Differences: Lower Limit	95% CI of Group Differences: Upper Limit	<i>p</i> ^b
Perphenazine	69.01					
Olanzapine	66.69	2.32	0.96	0.05	4.58	*
Quetiapine	69.38	-0.37	0.98	-2.69	1.95	
Risperidone	71.46	-2.45	0.98	-4.75	-0.14	*
		Months 9 and 12 (<i>n</i> = 731 observations from 401 patients)				
Perphenazine	61.28					
Olanzapine	62.21	-0.93	1.69	-5.43	2.68	
Quetiapine	64.24	-2.96	1.83	-7.86	0.99	
Risperidone	65.11	-3.83	1.78	-8.36	-0.23	*
		Months 15 and 18 (<i>n</i> = 553 observations from 307 patients)				
Perphenazine	60.29					
Olanzapine	59.67	0.62	1.88	-3.70	5.46	
Quetiapine	64.00	-3.71	2.15	-9.11	1.38	
Risperidone	60.23	0.06	2.04	-4.55	5.40	

^a Patients with tardive dyskinesia at baseline were excluded for all groups in this randomization. Because of missing data, two cases were excluded from these analyses.

^b Significance of paired differences between perphenazine and each SGA (**p* < 0.017 due to Bonferroni's correction for three comparisons).

LS indicates least squares.