

Should the PANSS Be Rescaled?

Michael Obermeier², Andreas Mayr^{1,2},
Rebecca Schennach-Wolff², Florian Seemüller²,
Hans-Jürgen Möller², and Michael Riedel²

²Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University Munich, Nussbaumstrasse 7, 80336 Munich, Germany

The design of the Positive and Negative Syndrome Scale (PANSS) with item levels ranging from 1 to 7 leads to the trivial result that the 30-item scale's zero level (no symptoms) is 30. This causes serious problems when ratios are calculated which always implicitly depend on a natural zero point (equals 0). Recent publications concerning efficacy of antipsychotics correctly suggest a subtraction of 30 points to every PANSS before calculating percent change (PC). Nevertheless, the traditional approach using uncorrected scores is still in common practice. This analysis aims to clarify which approach is the most appropriate from a statistical perspective. For analysis, data from a naturalistic study on 400 patients with a schizophrenic spectrum disorder and simulated data sets were used. While calculations concerning absolute score values and their differences are not affected, considerable problems arise in calculations of PC and related response criteria. Even significance levels of estimated treatment effects change, depending on the structure of the data (eg, baseline symptom severity). Using a PANSS version with items ranging from 0 to 6 would avoid such often neglected pitfalls.

Key words: scale level/minimum subtraction/percent change/simulation study

Introduction

The Positive and Negative Syndrome Scale (PANSS^{1,2}) is one of the most common scales in clinical studies for measuring symptom severity in patients with schizophrenia. Treatment effects relating the posttreatment score (PANSS₉₉) with the corresponding baseline measurement (PANSS₀) can be analyzed and compared. Various effect

measures have been discussed in statistical literature: Törnqvist et al³ compare up to 10 ways of measuring a relative difference resulting in the proposal of the log chance and the log percentage, while Berry and Ayers⁴ showed the high power of symmetrized percent change (PC) in statistical analyses. In the present article, we focus on the ordinary PC $100 \times (\text{PANSS}_{99} - \text{PANSS}_0) / \text{PANSS}_0$ because it is commonly used in schizophrenia research⁵ to indicate treatment effects: Response is typically defined as a distinct reduction level in terms of PC in the total score which has to be reached (eg, see Leucht et al,⁶ Marder and Meibach,⁷ Peuskens⁸). But regardless of which of the above-mentioned measures is used, its proper calculation confronts researchers with a severe pitfall.

The PANSS is an interval scale where calculating ratios is not appropriate due to the lack of a natural zero point. The item level of the 30 items ranges from 1 to 7, with 1 equaling “no symptoms,” resulting in a total score of 30 points for a patient with no symptoms. Hence before calculating ratios, the scale level has to be changed into a ratio scale by subtracting 30 points.

Unfortunately, this problem is often overlooked, and therefore, different calculation methods exist: While in some studies a general subtraction of 30 points has been applied (eg, Labelle et al^{9,10}), others obviously used the raw score (eg, Lee and Kim,¹¹ Sacchetti et al,¹² Food and Drug Administration¹³) or at least do not provide information as to whether the subtraction was carried out or not (eg, Spina et al,¹⁴ Honer et al,¹⁵ Breier et al,¹⁶ Kane et al¹⁷).

Because the different calculation methods might generate different significance levels, finally resulting in misinterpretations of treatment effects, there is a strong need for clarification on this subject.

Leucht et al^{6,18} have already emphasized the necessity of the 30-point subtraction for the calculation of PC. However, up to now, to the best of our knowledge, no systematic analysis has been performed to evaluate the impact of the different usage of PANSS on the results of schizophrenia studies.

Our aims were therefore (1) to clarify for which statistical procedure it is necessary to subtract the minimum of 30 points and (2) to investigate the effect on study results if the subtraction was omitted. Specifically, we focused on conditions which might lead to different results

¹To whom correspondence should be addressed; tel: +49-89-5160-5753, fax: +49-89-5160-5774, e-mail: andreas.mayr@med.uni-muenchen.de.

concerning significant group effects (eg, treatment effects), depending on the calculation method used (subtracting or not subtracting 30 points). Hence, we analyzed test decisions with and without subtraction in (1) a real data set of a naturalistic follow-up study and (2) in simulated data.

Patients and Methods

The Database

1. The real data included 400 patients with schizophrenia spectrum disorder (226 male and 174 female) treated under naturalistic conditions. Study protocol, main results, and specific study aims were described in detail elsewhere.¹⁷ The mean age was 35.5 ± 11.1 (mean \pm SD) years.
2. To generalize results and to allow detailed analysis of structural aspects, simulated data sets were included representing typical data of clinical group trials.

Statistical Analysis

We compared PC and response rates of the real data set between both calculation methods. In a further step, we compared test results between both procedures for group differences regarding percentage of PANSS reduction. For this purpose, we used linear models with the grouping variable as independent variable, focusing on the values of the test statistics (Wald tests).

Simulated data sets represented results of clinical trials and therefore contain simulated PANSS total at baseline (PANSS₀) and end point (PANSS₉₉), respectively. These data were produced for 2 assumed groups A and B (representing, eg, placebo vs verum) each including 500 patients. For generating simulated baseline data PANSS₀, we used a discrete parametric distribution which is geared to the empirical distribution of the real data sets.

To get an impression of a typical treatment course, we fit a linear model of PANSS₉₉ on PANSS₀ for the real data set. The estimates of this model were used to generate PANSS₉₉ data for the 2 different subgroups on the basis of the simulated baseline data. As with this procedure, PANSS₉₉ and PANSS₀ would be perfectly correlated ($\text{cor} = 1$); additionally, a Gaussian noise (data from a normal distribution with $\mu = 0$ and a certain σ) was added on PANSS₉₉ to reach a correlation structure comparable to the real data. The greater the σ , the weaker is the correlation between PANSS₉₉ and PANSS₀ and vice versa. To consider different scenarios, one parameter of the admission-distribution varied, while all other parameters remained fixed. For each combination of distribution parameters, we computed 100 different data sets and calculated the same statistical measures as for

Table 1. Real Data set; Group Effect Concerning PC?

	Mean PC Male (%) ^a	Mean PC Female (%) ^a	<i>t</i> Value ^b	<i>P</i> Value ^b
30 not subtracted	25.56	26.97	-1.82	0.07
30 subtracted	44.38	49.18	-1.69	0.09

Note: PC, percent change.

^aMean PC in male/female group from baseline to end point.

^bTest statistic and *P* value (Wald test) of the estimated group effect (male/female) on PC in a linear model.

the real data in each. Accordingly, we averaged over all data sets with the same parameter combination.

All analyses were performed using the statistical computing environment R 2.8.1.¹⁹

Results

Real Data

The real data set consisted of 400 patients treated under naturalistic conditions with a mean PANSS total at baseline of 71.17 ± 19.14 (mean \pm SD). To demonstrate the effect of different calculation methods on a test decision, we arbitrarily chose the grouping variable “gender.”

The results presented in table 1 address gender effects on the treatment course in a naturalistic design. In this example, the 2 methods obviously lead to different values of PC, but statistical testing still revealed the same results concerning the group effect.

Further on, we classified patients as treatment responders if they reached a specific reduction level from baseline on PANSS total score in terms of PC (20% or 50% reduction). Table 2 shows *z* and *P* values of logistic regression models, analogue to *t* values in the Gaussian linear model above.

In this example, the significance changes between the 2 methods in 1 case: The statistical testing of a possible gender effect using a 20% response criterion leads to contradictory results due to the different calculation methods.

The influence of the calculation method on PC is further illustrated in Figure 1. For each individual patient,

Table 2. Real Data Set; Group (Gender) Effect Concerning Response?

	20% Response		50% Response	
	<i>z</i> Value	<i>P</i> Value ^a	<i>z</i> Value ^a	<i>P</i> Value ^a
30 not subtracted	1.64	0.10	-0.05	0.96
30 subtracted	2.03	0.04	1.80	0.07

^aTest statistic and *P* value (Wald test) of the estimated group effect (male/female) on response in a logistic model.

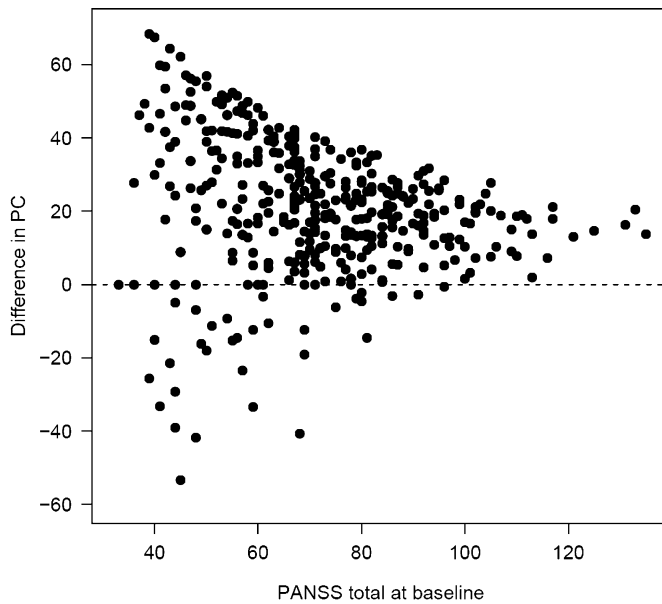


Fig. 1. Absolute Differences in Percent Changes Between Calculation Methods Depending on the Baseline Level in a Real Data Set. With decreasing baseline level, the differences between the calculation methods increase.

the difference in PC between the 2 methods is plotted against the baseline score. Depending on the calculation method, differences in PC increase with decreasing baseline level. Hence, a data set with many patients with a low PANSS at baseline will be more affected than a data set, where patients have higher scores.

Simulation Study

We modeled our simulated data on the previously considered real data set. With respect to the distribution of the PANSS₀ baseline data, a right skewed (discretized) gamma distribution was most similar to the real data set. The relationship between PANSS₀ and PANSS₉₉

was established using the parameters of a linear model on the real data set (an effect between group A and group B was produced by applying different slope parameters). The Gaussian noise, added to PANSS₉₉, had a σ of 15 and resulted in correlations between PANSS₀ and PANSS₉₉ from 0.39 to 0.59.

Table 3 shows some representative results regarding PC in the simulation study in relation to different levels of PANSS₀ and in combination with an existing vs a non-existing effect between groups A and B.

For the same 4 data sets, table 4 shows the corresponding results for the dichotomous outcome, with levels of 20% and 50% for response.

Considering test decisions in simulation studies without real group effect, both methods show the expected results: Mean t values are close to 0, which is far away from statistical significance. Nevertheless, the SD of the t value differences between the 2 methods clearly increases with decreasing baseline level indicating possible inconsistencies. When there is a true group effect, differences occur especially with low baseline levels. Regarding PC, the method with subtraction seems to be more conservative; however, there were also data where this method showed a higher (absolute) t value.

With regard to responder analyses, it is conspicuous that with increasing response level and decreasing baseline level without subtraction of 30 points, the number of responders is reduced. Although the z values are quite consistent in studies where no real group effect exists, results differ clearly for the most other data sets: Without subtraction, the strong response criterion leads, apart from the very low responder rates, also to lower (absolute) z values, showing lower significance for the grouping variable.

The last column of each table shows the percentage of simulated studies in which both methods lead to different conclusions regarding significance. Depending on the baseline level and the analyzed outcome criteria, the

Table 3. Simulation Study; Group Comparison Between A and B With Respect to PC

ID ^a	Effect	Mean PANSS ₀ ^b		t Values ^c		Method Differences	
		Group A	Group B	30 Not Subtracted	30 Subtracted	SD (t Difference) ^d	Significant Change ^e
1	No	62.59	62.43	0.14	0.15	0.60	3
2	No	72.53	72.55	0.07	0.03	0.38	1
3	Yes	62.59	62.43	-2.41	-1.44	0.60	40
4	Yes	72.53	72.55	-2.91	-2.25	0.39	17

Note: PANSS, Positive and Negative Syndrome Scale; PC, percent change.

^aID of simulation study.

^bPANSS total: mean at baseline.

^cMean t value (Wald tests) of the estimated group effect on PC in a linear model.

^dEmpirical SD of differences in t values between both methods ($SD(t_1 - t_2)$).

^eNumber of data sets where the results (Wald tests) differ regarding significance (one method: significant effect found, second method: no effect found; number of data sets each time: 100).

Table 4. Simulation Study; Group Comparison Between A and B With Respect to Dichotomous Response

ID ^a	<i>z</i> Values ^b		Rates ^c		Significant Change ^d
	30 Not Subtracted	30 Subtracted	30 Not Subtracted	30 Subtracted	
Group comparison between A and B with respect to 20% response					
1	-0.05	-0.09	50	66	3
2	0.00	0.12	62	76	6
3	-1.93	-1.88	47	63	28
4	-2.25	-2.05	59	73	26
Group comparison between A and B with respect to 50% response					
1	0.06	0.01	6	41	8
2	0.15	0.08	10	48	11
3	-1.32	-2.17	5	38	51
4	-2.04	-2.15	9	44	32

^aID of simulation study.

^bMean *z* value (Wald tests) of the estimated group effect on PC in a logistic model.

^cMean responder rates.

^dNumber of datasets where the results (Wald tests) differ regarding significance (one method: significant effect found, second method: no effect found; number of data sets each time: 100).

number of studies with inconsistent test decisions can rise to above 50%.

Discussion

Theoretical Implications

Statistics which refer to absolute values of the PANSS are not affected, regardless of whether 30 points were subtracted or not. By contrast, differences between the 2 scale levels appear when ratios are calculated, as in response analyses. A simple numerical example might demonstrate this: Without subtraction, a 50% reduction of a PANSS baseline level of 50 would result in a score of 25, which is impossible given the minimum of 30. Furthermore, a 100% reduction is rendered impossible. On the other hand, the disappearance of all symptoms leads to a PC of $\frac{30-50}{50} \times 100 = -40\%$, which does not reflect that the patient is asymptomatic.

Subtracting 30 points from the PANSS equals a score with items ranging from 0 to 6 instead of ranging from 1 to 7. This leads to a change in the PANSS level of measurement: Because there is no natural zero point for the 1–7 version, the PANSS in its original version is an “interval scale” on which ratio operations such as calculating proportions are not suitable,²⁰ as seen in the above example. The subtraction changes the level of measurement into a “ratio scale” by constructing the zero point.

Using the unchanged interval scale means underestimating PC (in both directions: $|PC_{\text{interval}}| \leq |PC_{\text{ratio}}|$), which leads to the conclusion that the correct calculation of the PC results in more patients fulfilling response criteria (see tables 1 and 4). Additionally, it results in different test statistics (and therefore *P* values) of statistical hypothesis tests for group differences, eg, differences between medications, as shown in this study.

Besides the obvious inequality of the 2 procedures, quantifying the effect of a wrong calculation is less trivial. In this context, the question arises as to which one is more likely to reveal a significant difference between treatment groups. Unfortunately, a general result (\leq or \geq) can hardly be obtained because the relation between both calculation methods follows a nonlinear function. Nevertheless, according to our simulations, the following points influencing the statistical outcome have to be considered:

1. Location and variance of PANSS₀ influence the difference between results of both calculation methods: The higher the PANSS₀, the smaller is the slope of the nonlinear function mentioned. Therefore, with decreasing level of PANSS₀ as well as with increasing variance, which causes a greater number of lower values, the difference between calculation methods as well as its variance will increase (see figure 1; tables 3 and 4).
2. Concerning the dichotomous outcome “response,” which is usually defined in terms of a special level of PC (20%, 30%, ...), subtracting 30 points leads to more patients reaching the response level (table 4). Apart from this, there is a further important theoretical aspect.

Using the interval version of the scale, a higher response level leads to more patients who are not able to become responders at all: With a 20% criterion, it is impossible for patients with an admission score of 37 or lower to become responders. At a response level of 50%, a baseline score of 59 already precludes a patient from fulfilling the criteria, which probably affects a reasonable number of patients. In other words, this approach indirectly excludes a significant number of

patients a priori from end point analysis who might otherwise have fulfilled the criterion.

Implications for Researchers and Clinicians

Results of a study in which PCs were calculated without a 30-point subtraction (1–7 scale) might be quite different compared with the (correct) calculation based on the ratio (0–6) scale, even regarding significance. Considering the 20% response criterion for the presented real data set, the correct analysis leads to the conclusion that there is a significant group effect, while an analysis based on the 1–7 scale leads to the opposite result (see table 2). The results of the simulation study show in some situations a rate of more than 50% of inconsistent test decisions (see table 4).

Unfortunately, due to the nonlinearity of the problem, data provided in standard publications of medication trials are often not sufficient to estimate whether or not results were affected by the PC calculation method, and if so, in which direction.

This issue might have concrete and far-reaching implications as in drug approvals. For example, in some recent published approval studies of atypical antipsychotics, it was not clearly stated which method was chosen.^{16,17} In at least one, it appears very likely that the wrong procedure might have been used.¹³ This example illustrates the high relevance of an international consensus on the implementation of this issue.

The most straightforward approach with a minimum source of errors would be a rescaling of the PANSS from 0 to 6. To avoid the possibility of new uncertainty, the 0–6 scale could be referred to as “PANSS (ratio version).” Using this, little add-on should prevent confounding results from the 2 PANSS versions. At first glance, this suggestion may sound extreme, but 2 existing PANSS versions which are clearly separated by their denotation will be less confusing and prone to errors than a scale which forces the researcher to transform it before calculating PCs and the reader to guess if this transformation was made or not. Therefore, this solution might help in avoiding further confusion in the work of schizophrenia researchers as well as in daily clinical usage.

However, the introduction of a new version (change of the user manuals, new publication, and new printing) would cause considerable efforts and might be not very feasible. An alternative could be the subtraction of the respective possible minimum prior to any PC analysis. However, this would implicate that for all PC-related calculations, eg, the calculation of PC for PANSS subscores, the correct minimum, depending on the amount of subscore items needs to be considered. In addition, a correct description of when and where the subtracted PANSS scores were used and where they were not would be essential. This in turn bears considerable risks for errors.

Further discussions appear to be necessary to reach a broad consensus in the psychiatric community on future work with the PANSS. Until this consensus is found, at least a clear declaration of how the PANSS was used should be stated in each publication.

Acknowledgments

The real data study was conducted at 14 psychiatric hospitals: Aachen (P. Hoff, K. Podoll), Augsburg (M. Schmauß, T. Messer, M. Eichinger), Berlin (I. Heuser, M. Jockers-Scherübl), Bonn (W. Maier, K.-U. Kühn, M.R. Lemke, R. Hurlmann, W.P. Hornung, E. Rosen), Cologne (J. Klosterkötter, W. Huff), Düsseldorf (W. Gaebel, A. Klimke, M. Eickhoff, M. von Wilmsdorff), Essen (M. Gastpar, V. Reißner), Gabersee (G. Laux, B. Hermann, B. Plichta), Göttingen (E. Rütther, D. Degner), Haar (H. Pfeiffer, M. Albus, S. Scharf-Büssing), Hamburg (D. Naber, D. Golks), Mainz (L.G. Schmidt, B. Kaufmann-Grebe), Munich (H.-J. Möller, R. Bottlender, M. Riedel, M. Jäger, C. Schorr, B. Schillinger, C. Mirlach), and Tübingen (G. Buchkremer, M. Mayenberger). We would like to thank T. Coutts for the linguistic revision of the manuscript.

References

1. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr. Bull.* 1987;13:261–276.
2. Kay SR, Opler LA, Fiszbein A. *The Positive and Negative Syndrome Scale (PANSS) Manual*. Toronto, ON: Multi-Health Systems Inc.; 2000.
3. Tornqvist L, Vartia P, Vartia YO. How should relative changes be measured? *Am Stat.* 1985;39:43–46.
4. Berry DA, Ayers GD. Symmetrized percent change for treatment comparisons. *Am Stat.* 2006;60:27–31.
5. Fleischhacker WW, Kemmler G. The clinical relevance of percentage improvements on the PANSS score. *Neuropsychopharmacology.* 2007;32:2435–2436.
6. Leucht S, Davis JM, Engel RR, Kane JM, Wagenpfeil S. Defining ‘response’ in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology.* 2007;32:1903–1910.
7. Marder SR, Meibach RC. Risperidone in the treatment of schizophrenia. *Am J Psychiatry.* 1994;151:825–835.
8. Peuskens J. Risperidone in the treatment of patients with chronic schizophrenia: a multi-national, multi-centre, double-blind, parallel-group study versus haloperidol. Risperidone Study Group. *Br J Psychiatry.* 1995;166:712–726.
9. Labelle A, Boulay LJ, Lapierre YD. Retention rates in placebo- and nonplacebo-controlled clinical trials of schizophrenia. *Can J Psychiatry.* 1999;44:887–892.
10. Labelle A, Light M, Dunbar F. Risperidone treatment of outpatients with schizophrenia: no evidence of sex differences in treatment response. *Can J Psychiatry.* 2001;46:534–541.
11. Lee BH, Kim YK. Increased plasma brain-derived neurotropic factor, not nerve growth factor-Beta, in schizophrenia

- patients with better response to risperidone treatment. *Neuropsychobiology*. 2009;59:51–58.
12. Sacchetti E, Galluzzo A, Valsecchi P, Romeo F, Gorini B, Warrington L. Ziprasidone vs clozapine in schizophrenia patients refractory to multiple antipsychotic treatments: the MOZART study. *Schizophr Res*. 2009;110:80–89.
 13. FDA: U.S. Food and Drug Administration. Drug approval package for zyprexa intramuscular (olanzapine) injection, Application No. 021253, Approval Date 3/29/2004. http://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/21253_Zyprexa.TOC.cfm. Accessed July 13, 2005.
 14. Spina E, Avenoso A, Facciola G, et al. Relationship between plasma risperidone and 9-hydroxyrisperidone concentrations and clinical response in patients with schizophrenia. *Psychopharmacology (Berl)*. 2001;153:238–243.
 15. Honer WG, Thornton AE, Chen EY, et al. Clozapine alone versus clozapine and risperidone with refractory schizophrenia. *N Engl J Med*. 2006;354:472–482.
 16. Breier A, Meehan K, Birkett M, et al. A double-blind, placebo-controlled dose-response comparison of intramuscular olanzapine and haloperidol in the treatment of acute agitation in schizophrenia. *Arch Gen Psychiatry*. 2002;59:441–448.
 17. Kane JM, Carson WH, Saha AR, et al. Efficacy and safety of aripiprazole and haloperidol versus placebo in patients with schizophrenia and schizoaffective disorder. *J Clin Psychiatry*. 2002;63:763–771.
 18. Leucht S, Davis JM, Engel RR, Kissling W, Kane JM. Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. *Acta Psychiatr Scand Suppl*. 2009;438:7–14.
 19. R Development Core Team Organization: R Foundation for Statistical Computing. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <http://www.R-project.org>. 2009.
 20. Fahrmeir L, Künstler R, Pigeot I, Tutz G. *Statistik: Der Weg zur Datenanalyse, 4*. Berlin, Germany: Springer; 2003.