Aus der Klinik für Psychiatrie und Psychotherapie
der Ludwig-Maximilians Universität München
Direktor: Prof. Dr. Hans-Jürgen Möller

The Positive and Negative Syndrome Scale for Schizophrenia:
An Established Rating Instrument in Need of Clarification

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Michael Johann Obermeier

aus
Deggendorf

im Jahr
2012

Berichterstatter:	Prof. Dr. Dr. h.c. mult. Hans-Jürgen Möller
Mitberichterstatter:	Prof. Dr. Thomas Bronisch
	Prof. Dr. Dr. Wilfried Günther
Mitbetreuung durch den	
promovierten Mitarbeiter:	Prof. Dr. Rolf Engel
Dekan:	Prof. Dr. med. Dr. h.c. Maximilian Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 24. 09. 2012

Table of contents

Abstract

The modern debate about schizophrenia began over 100 years ago, with Kraepelin's description of "dementia praecox". Despite this, central aspects of the disease remain mysterious and the disease itself is still associated with a high probability of an enduring limitation of the patient's quality of life. While several conceptions of schizophrenia exist and are still under discussion, at least a provisional consensus regarding a valid measure of schizophrenia seems to have been reached: The Positive and Negative Syndrome Scale (PANSS) quantifies the current state of a person with schizophrenia by combining 30 different schizophrenia-associated symptoms into a single scale value.

Even though the scale is widely used and is the measure of choice in many clinical trials, its psychometric properties are still the reason for serious confusion. In many research papers, one important fact about the PANSS is overlooked: it is an interval scale and, therefore, straightforward calculations of proportions are not appropriate. In other words, calculating simple percentage changes is incorrect and a prior scale correction is required. These kinds of calculations often appear in conjunction with responder analyses, as the definition of response is usually based on a predefined cut-off in terms of percent scale change. Two of the presented papers of this thesis are dealing with this urgent problem: using real data as well as simulated data sets, it is shown that ignoring the scale level of the PANSS can, in many cases, even lead to false test decisions concerning an examined treatment effect. Furthermore, an analysis of the problem's urgency with regard to academic discussions, performed by way of a systematic study of literature in the highest-ranked journals dealing with schizophrenia, showed that incorrect calculations are widespread in the literature and that there is a strong need for a general clarification.

As incorrectly calculated percent changes might be a reason for the published low cut-offs of response, as e.g. 20% or 30% cut-offs, the third included article in this thesis analyzes the association of correctly calculated percent changes in the PANSS with a generally measured therapy response. An equipercentile linking of percent PANSS changes and the improvement item of the Clinical Global Impression Scale (CGI) confirmed the choice of a considerably higher response cut-off of 50%.

The combined conclusion of the three included articles is the emphasis on the need for a general methodological consensus in schizophrenia research. Valid and replicable research is only possible on the basis of generally accepted methods that rely on the correct application of scale theory in these studies.

Zusammenfassung

Vor mehr als 100 Jahren fiel mit Kraepelins Beschreibung der sogenannten „Dementia Praecox" der Startschuss für die moderne Schizophreniedebatte. Trotz dieser langen Zeitspanne bleibt diese Erkrankung in zentralen Punkten unverstanden und ist immer noch mit einer hohen Wahrscheinlichkeit dauerhafter Einschränkungen der Lebensqualität Erkrankter verbunden. Während aber auf der einen Seite verschiedene Schizophreniekonzepte nebeneinander existieren und diskutiert werden scheint es inzwischen zumindest einen vorläufigen Konsens über die Beantwortung der Frage zu geben, wie die Ausprägung der Erkrankung zu messen ist: Die „Positive and Negative Syndrome Scale" (PANSS) quantifiziert den aktuellen Zustand eines schizophrenen Patienten durch das Zusammenfassen von 30 verschiedenen, mit Schizophrenie assoziierten Symptomen zu einem Skalenwert. Aber obwohl die Skala weit verbreitet und das Messinstrument der Wahl in vielen klinischen Studien ist, sind ihre psychometrischen Eigenschaften immer noch der Grund für schwerwiegende Konfusionen. In vielen wissenschaftlichen Artikeln wird eine wichtige Eigenschaft der PANSS übersehen: sie ist eine Intervallskala und als solche ungeeignet für die Berechnung von Proportionen. Mit anderen Worten: das simple Berechnen von prozentualen Veränderungen ist falsch, es bedarf hier einer zusätzlichen Korrektur. Diese prozentualen Veränderungen tauchen aber ihrerseits häufig im Zusammenhang mit Responder-Analysen auf, da die Definition von Response üblicherweise über eine a priori definierte prozentuale Skalenverbesserung erfolgt. Zwei der in dieser Dissertation dargestellten Artikel befassen sich direkt mit diesem gravierenden Problem: mithilfe von sowohl realen als auch simulierten Daten wird gezeigt, dass das Ignorieren des Skalenlevels der PANSS in vielen Fällen sogar zu falschen Testentscheidungen bezüglich eines vermuteten Behandlungseffekts führen kann. Eine anschließende systematische Literatursuche in den höchstrangigen Schizophrenie-Journalen zeigt, dass der Berechnungsfehler in der Literatur weitverbreitet ist, es also generell großen Klärungsbedarf bei diesem Thema zu geben scheint.

Falsch berechnete prozentuale Veränderungen ihrerseits könnten mit ein Grund sein für gebräuchliche, niedrige Schwellenwerte für Response, wie beispielsweise 20%- oder 30%-Kriterien, weshalb sich der dritte Artikel dieser Arbeit mit dem Zusammenhang von, richtig berechneten, prozentualen Veränderungen der PANSS

mit einem allgemein gemessenen Therapieresponse befasst. Eine equiperzentil Linking Analyse von prozentualen PANSS-Veränderungen mit dem Verbesserungs-Item der „Clinical Global Impression Scale" (CGI) bestätigt die Wahl eines wesentlich höheren Schwellenwertes von 50%.

Zusammengefasst betonen die drei dargestellten Artikel die Notwendigkeit eines generellen methodologischen Konsensus in der Schizophrenieforschung. Valide und reliable Forschung ist nur möglich auf der Basis generell akzeptierter Methoden, die sich ihrerseits auf die korrekte Anwendung skalentheoretischer Erkenntnisse stützen.

Introduction

Schizophrenia is a severe mental disease that is characterized by different psychopathological symptoms like disturbance of the affect, difficulty thinking or dysfunction of perception. Its lifetime prevalence is estimated at around 1% [31]. Even though there has been an ongoing debate about schizophrenia for over 100 years, its aetiology, course, and treatment are still not completely understood.

Several different theories offer explanatory approaches for the emergence of schizophrenic symptoms. Twin studies have shown that the risk of illness increases with closer biological relationship to diseased persons, which suggests a genetic predisposition to schizophrenia. Further explanatory approaches consider abnormal dopamine activities to be an important factor for this disease [13], others describe early viral or bacterial infections [9] and also perinatal complications [10] as potential risk factors. Modern treatment strategies refer to the dopamine approach since they are based on medical applications in terms of antipsychotics, which influence the dopamine metabolism. Despite these treatment regimens, the course of schizophrenia is heterogeneous, with some patients still suffering from a poor prognosis and a high probability of enduring limitations in their quality of life.

One reason for this might be the heterogeneity of its pathophysiological underpinnings and clinical characteristics, ranging from blunted affect and social withdrawal to disturbances in complex thinking to the highly conspicuous symptoms of delusions or hallucinations. This wide variety of symptoms might have contributed to the initial difficulties of an adequate definition of this illness. Kraepelin offered a first description in 1899 [21] by combining several typical courses of the disease, beginning in adolescence with the onset of intellectual decline and ending with an early stage of dementia leading into an entity disease called "dementia praecox" (early dementia). This definition included an essential early start of the disease as well as a mental deterioration, which, as we now know, are not necessarily both present. Nevertheless, it is to Kraepelin's merit that these symptoms were distinguished from affective disorders or, in Kraepelin's words, from manic-depressive illness. As it was clear that Kraepelin's conception was not wide enough, his first attempt of describing this mental illness was replaced by a new definition. In 1911, Bleuler published his monograph "Dementia praecox oder Gruppe der Schizophrenien," [7] where he broadened the definition by describing

basic symptoms (affect, association, ambivalence, autism and others) and several accessory symptoms (like hallucinations, delusion or catatonia). He introduced the term "schizophrenia" as a description of a group of mental diseases with these symptoms.

Schneider [40] added a new point of view on schizophrenia by introducing the differentiation between first- and second-rank symptoms. In particular, his careful perception of psychiatric diagnoses still influences the debate: if aetiology and pathogenesis of a disease remain unknown, as is the case in schizophrenia, he claimed that one cannot talk about differential diagnoses but only about differential typologies.

This point of view contributed to the modern operationalized diagnostics used in ICD-10 and DSM-IV. Here, schizophrenia is classified by a long list of phenotypical symptoms including the symptoms described by Bleuler as well as those described by Schneider. Both ICD-10 and DSM-IV classify schizophrenia into diverse subtypes such as the paranoid or the catatonic subtype by combining special groups of symptoms. This classification is solely descriptive without any prognostic capacity.

This prognostic capacity is assigned to the conception from Crow [12] and his distinction between type-I and type-II Schizophrenia. The basis of this classification is the distinction between positive and negative symptoms. Positive symptoms are symptoms with productive character such as hallucinations, delusions or bizarre behaviour and are supposed to be dominant in type-I. Type-II is dominated by negative symptoms. These symptoms are characterized by the absence of normal experiences and appear, for example, in deficits of the affect, of thinking and communication or in decreased motivation. This classification allows a connection between diagnosis and prognosis, as negative symptoms are supposed to be associated with poorer response to antipsychotics: type-I is therefore called acute schizophrenia, type-II chronic schizophrenia. Although this simplifying concept of type-I and type-II schizophrenia could not be validated [31], the distinction in positive and negative symptoms plays a central role in modern discussion and is part of important psychiatric rating scales.

These scales are of essential interest in schizophrenia research: in psychiatric research in general, no biological markers exist offering an objective measure of disease severity. Therefore, other measures are needed to obtain as precise of an impression of a patient's condition as possible. By describing the intensity of

relevant symptoms with a combining scale, value rating scales offer a good insight into a patient's current situation with regard to his symptoms and allow for comparison both between different patients and different time periods for a single patient. Comparing the scale value courses between differently treated patient samples allows a substantiated efficacy assessment with regard to the applied treatment strategies.

Particularly in approval studies of new medications, these assessments are of essential importance as they are the basis for the approval of a new treatment: only medications showing significant positive effects according to the applied psychiatric rating scales are supposed to offer a benefit to patients. A central requirement in this context is a satisfying external validity, which means the assumption that the applied scales adequately measure the patients' statuses: different score values should represent patients' states that are actually different and a score-change in time should correctly indicate a shift in a single patient's status. The condition for this requirement is generally a solid and comprehensible scale validation based on classical test or modern item response theory and, especially, an extensive rater training yielding a correct and comparable scale application by individual raters.

Examples of common rating scales used in schizophrenia are scales measuring the global functioning of patients like the Global Assessment Scale of Functioning (GAF, [1]) or those measuring a patient's quality of life such as the Lancashire Quality of Life Profile (LQLP, [43]). The first scales measuring particularly positive and negative symptoms were the Scale for the Assessment of Negative Symptoms [2] and the Scale for the Assessment of Positive Symptoms [3]. In 1987, Kay et al. developed the Positive and Negative Syndrome Scale [17] by combining two established rating systems, the Brief Psychiatric Rating Scale (BPRS, [36]) and the Psychopathology Rating Schedule (PRS, [41]) into a single rating instrument. Their aim was to develop a standardized scale, which measures both positive and negative symptoms with the same priority, reacts sensitively to drug-related changes, and includes a measure of general psychopathology. Their development has become a complete success and this scale is likely used the most in current schizophrenia research.

The PANSS consists of 30 items measuring specific symptoms, each item ranging from 1 (absent) to 7 (extreme). A total score is built by simply adding up the single items. These psychometric properties are the background for this current research.

The scale is based on a formalized psychiatric interview taking approximately 45 minutes and requiring an accurate rater training to reach a satisfying level of reliability. A detailed manual [16,18] offers a broad description of the aim of this instrument and of the interview procedure, including information about its beneficial psychometrical properties.

As indicated by the name, a special focus lays on the measurement of positive and negative symptoms: there are seven items measuring positive symptoms, seven for negative symptoms and 16 items corresponding to general symptoms. The positive symptoms comprise of delusions, conceptual disorganization, hallucinations, hyperactivity, grandiosity, suspiciousness and hostility and together form a positive subscale. In the negative items, the symptoms blunted affect, emotional withdrawal, poor rapport, apathetic social withdrawal, difficulty in abstract thinking, lack of spontaneity and stereotyped thinking are included. Together, they form the negative subscale. A bipolar composite scale can be formed by subtracting the negative from the positive score. The 16 general and respectively global items measure symptoms like anxiety, tension, mannerism, unusual thought contents or disorientation.

Since its initiation, several post validity studies using factor analytical methods empirically identified a five-factor structure of the scale (e.g. [6], [28], [26]) including a negative factor, a positive factor, a disorganized factor, an anxiety or depression factor and an excitement factor.

The PANSS total as well as its subscales and factors serve as a measurement of the patient's current symptom status. While, for example, the PANSS positive subscale tells something about the level of psychotic symptoms, the PANSS total score is supposed to measure the patient's condition with respect to the illness in general.

Accordingly, the PANSS is used in clinical studies to describe the psychopathological condition of a patient or to quantify effects of special treatments. Several possibilities for the quantification of a treatment effect are established: unadjusted before-after differences in scale values can be used as a measure of effect, as well as proportional values, where this difference has been adjusted to the scale value before treatment. Dichotomous outcomes are also established: they define an aim in terms of a PANSS measure that describes a successful treatment. A common outcome domain are the widely-used consensus criteria for remission by Andreasen et al. [4], which describe a remitted status of a patient's schizophrenia disorder by focussing on eight PANSS items measuring core

symptoms of schizophrenia. Another important outcome definition is the response criterion, which is based on the percent change in the scale over time. Patients who reach at least a pre-specified proportional PANSS-change during treatment are classified as responders. The cut-off for this criterion is often set to 50%, but other levels are also common.

Obviously, percent changes (PC) play an important role in PANSS-measured schizophrenia studies, either as continuous effect-measures or as a basis for the classification of treatment responders. In this context, unfortunately, one important characteristic of the scale is often overlooked. This point is the topic of the first paper presented in this thesis [34] and is described in the following.

Because all items of the scale range from one to seven, the minimum possible value of the total score is 30 for patients with no symptoms at all. This artificial base-level of 30 points leads to the fact that the PANSS is an interval scale, where straightforward calculations of ratios are not appropriate, meaning that a score reduction from 80 to 40 points does not represent a 50% reduction! In other words, if, for example, the PANSS score of a patient is 50 points at baseline and 30 points at endpoint the patient has obviously responded completely, as he has lost all relevant (i.e. PANSS-measured) symptoms. However, with a naïve PC calculation ($\frac{(50-30)}{50}=0.4$) he would only reach a percentage improvement of 40%, not even fulfilling the 50% response cut-off.

The reason for this gap lays in the absence of a natural zero-point of the scale, which is also the key for a solution of this problem: subtracting the potential minimum of 30 points lets the PANSS start at zero and changes the scale level from an interval to a ratio scale. With this rescaled PANSS, calculations of PC are appropriate and lead in the described example to a 100% reduction ($\frac{(50-30)}{(50-30)}=1$).

It is easy to see that results not based on proportions are not affected at all by this rescaling procedure, as absolute differences stay the same regardless of the possible scale minimum.

Based on this score characteristic and the fact that it is often overlooked, the paper tries to find an answer to the question of the relevance of correct PC calculations: is it possible that study results differ according to their PC calculation method? And if so, in which direction do they differ? To find an answer, real data from a naturalistic trial by the German research network on schizophrenia [44] have been used as well

as data from a simulation study. In each study, both ways of calculating PC have been performed and their results concerning a test on group difference have been compared.

Indeed, serious differences have been found in the real as well as in the simulated data with, in part, over 50% difference in test decisions, i.e. tests where a significant effect was found with one method while the same data and the same test using differently-calculated PC values yielded a non-significant effect with the other method.

The central point of this paper is the fact that results of both methods, with or without rescaling, are, strictly speaking, incomparable with each other and that a consensus in the psychiatric community about a solution to this problem is needed.

Indeed, several solutions seem possible and are discussed in the article. In a comment to this paper, Leucht et al. [25] proposed the radical solution of a strictly rescaled version of the PANSS with each item going from zero to six instead of one to seven. However, the probability of success for this proposal depends on the awareness of the problem in the psychiatric community. While the paper finds an answer to the question about the possible impact of incorrect PC calculation on the result of a study, the dimension of this problem in general, i.e., its prevalence in the literature remains unclear. It may be negligible if only a small minority of authors use uncorrected score values while the larger majority rescales the PANSS before calculating PCs. On the other hand, a widespread prevalence of incorrectly-computed PCs would mean that a good portion of published results need to be recalculated before they can be compared to other results.

In the first paper of this thesis, some examples of articles with obviously incorrect calculations are shown, even including an application study, in addition to a number of papers with correct calculation. These examples were all found by a more or less unstructured literature study and therefore do not provide a representative profile of scientific literature using PANSS PCs. To get a more meaningful impression of the problem's prevalence, a systematic study of the literature is needed. This was the motivation for the next paper, which includes the results of such a study [35].

For this paper, articles from the 10 highest-ranked psychiatric journals, excluding those focusing on topics not related to the PANSS and schizophrenia, were found by a systematic literature search using the PRISMA statement [30]. The methods of PC calculation in the papers were analysed with regard to the presence or lack of

rescaling and these results associated to the prominence of PC in the respective articles.

The research in January 2011 ultimately confirmed the apprehension that most of the articles using PANSS PC values did not use the score correction. Even in articles with PANSS PC as a primary outcome, this outcome measure was frequently calculated incorrectly. The results of this literature study highlight the dimension of the problem and the urge for a general consensus in the psychiatric community about how to handle it.

Besides resulting in possibly incorrect test decisions, the incorrect or unclear PC calculation method means that two researchers might not be talking about the same thing if they discuss PANSS-measured outcomes. While one researcher observes high, correctly-calculated response rates using a 50%-criterion, another author calculates far lower response rates with the same criterion but without rescaling. In the comment on the prior paper, Leucht et al. [25] express the assumption that incorrect calculations might partly be the reason for low response rates and low cut-offs in studies.

Indeed, many different response cut-offs are established and widely used: a 20%-criterion is common (e.g. [37], [33]) as well as a 50% criterion (e.g. [38], [32]), and there are also examples for 30% (e.g. [19], [29]) and 40% criteria (e.g. [14]). Leucht et al. [22] proposed using a 50% cut-off for acutely-ill and a 25% cut-off for treatment-resistant populations. Furthermore, the PANSS reduction should be presented in a table offering the results of different cut-offs in 25%-steps to provide a comprehensive overview and to evaluate whether the effects were consistent. The European Medicines Agency (EMEA) accepts responder definitions above a 30%-criterion and recommends the presentation of additional results with alternative criteria for sensitivity analyses [11].

Nevertheless, it seems important to get an idea of what different cut-offs stand for. Again, Leucht et al. offered an initial answer to the question "What does the PANSS mean?" [24]. Using an equipercentile linking approach, they linked PANSS values to concurrently rated values of the Clinical Global Impression Scale [15], which describes a patient's overall clinical state. The CGI consists of two main subscales, each ranging from one (the best) to seven (the worst), which measure the patient's current state (CGI-severity) and the patient's state improvement since the beginning of a study (CGI-improvement). It is universally applicable for measuring the status of

depressed patients as well as the status of schizophrenic patients and is therefore, and because of its quick handling, a widely used rating instrument for measuring treatment effect.

Linking the PANSS and the CGI provides, in simple terms, a kind of translation of special PANSS (improvement-) values into the graduation of the CGI, which is intuitively easier to understand. The idea of equipercentile linking is to find the scores for two scales that correspond to each other with regard to their percentile rankings. The procedure is described in detail in Kolen and Brennan [20].

Applying equipercentile linking, Leucht et al. found a PANSS PC of around 50% being associated with a CGI-improvement of 2, which means "much improved" and would be a reasonable choice for a response criterion. In a replication analysis, Levine et al. [27] confirmed this association with a PANSS PC of 45-49% fitting to a CGI-I of 2.

While these linking analyses were based on data from several, international clinical trials, Schennach-Wolff et al. [39] aimed to replicate and validate the results in an own study using data from the aforementioned naturalistic trial of the German research network on schizophrenia [44]. This article is the third paper included in this thesis. It consists of three different linking analyses: one between PANSS total and CGI-severity, a second between PANSS PC and CGI-improvement and a third between PANSS absolute change again with CGI-improvement. All of these analyses were performed during different time periods.

While the results show a huge difference to the results from Leucht et al. and Levine et al. in the linking of PANSS total and CGI-severity, the analyses concerning PANSS PC and CGI-improvement are more similar: a CGI value of 2 (much improved) fits very well to a PANSS improvement of 50%. Linking the absolute PANSS change with the CGI-improvement, a moderately convex shape of the linking graphs appears, which may be a sign of the fact already described by Leucht et al. [23], that CGI-improvement measures relative change more than absolute change.

The most significant result of this paper with regard to the present thesis is the confirmation of a reasonable, high cut-off as response criterion: a 20% PANSS reduction refers to a CGI-improvement of 3, which only means "minimally improved" and therefore does not seem to be enough to classify a patient as a responder to a treatment. The correct choice for a cut-off in the sense of this linking

analysis would be a 50% reduction, fitting to a CGI-improvement of 2 ("much improved").

In summary, the necessity of appropriate definitions and methods in psychiatric research should continue to be emphasized. Moreover, even if it might be trivial to note, it is important to achieve a generally accepted consensus about what definitions and methods should be used. Discussions are needlessly complicated if two researchers think they are debating the same things, but indeed they are not. Therefore, a consensus must be reached about how to handle the PANSS in the future, i.e., if it should be rescaled and generally renamed, and also with regard to a universally valid response criterion for schizophrenic patients.

Contributions

The author of this thesis significantly contributed to all three included articles: for the first [34], he drafted the general idea and then performed all analyses including the simulation study. As lead author, he also wrote the first draft and was involved in each further step of the paper's development. The conception of the second paper [35] was also elaborated by the author, including the study of the literature and the review of the articles found. Again, the author wrote the first draft and was highly involved until the article's final version. For the third paper [39], the author performed all statistical analyses, including the description of the methods in the article. At each step of the paper, he contributed a critical revision of the entire manuscript.

Reference List

1. American Psychiatric Association: **Diagnostic and statistical manual of mental disorders, DSM-IV**, 4 edn. Washington DC: 1984.

2. Andreasen NC. **The Scale for the Assessment of Negative Symptoms (SANS).** 1983. Iowa City, Iowa, The University of Iowa.

3. Andreasen NC. **The Scale for the Assessment of Positive Symptoms (SAPS).** 1984. Iowa City, Iowa, The University of Iowa.

4. Andreasen NC, Carpenter WT, Jr., Kane JM, Lasser RA, Marder SR, Weinberger DR: **Remission in schizophrenia: proposed criteria and rationale for consensus.** *Am J Psychiatry* 2005, **162:** 441-449.

5. Beasley CM, Jr., Sanger T, Satterlee W, Tollefson G, Tran P, Hamilton S: **Olanzapine versus placebo: results of a double-blind, fixed-dose olanzapine trial.** *Psychopharmacology (Berl)* 1996, **124:** 159-167.

6. Bell MD, Lysaker PH, Beam-Goulet JL, Milstein RM, Lindenmayer JP: **Five-component model of schizophrenia: Assessing the factorial invariance of the positive and negative syndrome scale.** *Psychiatry Research* 1994, **52:** 295-303.

7. Bleuler E: **Dementia praecox oder Gruppe der Schizophrenien.** Deuticke, Leipzig - Wien: 1911.

8. Breier A, Berg PH, Thakore JH, Naber D, Gattaz WF, Cavazzoni P *et al.*: **Olanzapine versus ziprasidone: results of a 28-week double-blind study in patients with schizophrenia.** *Am J Psychiatry* 2005, **162:** 1879-1887.

9. Brown AS: **The risk for schizophrenia from childhood and adult infections.** *Am J Psychiatry* 2008, **165:** 7-10.

10. Cannon TD, Rosso IM, Hollister JM, Bearden CE, Sanchez LE, Hadley T: **A prospective cohort study of genetic and perinatal influences in the etiology of schizophrenia.** *Schizophr Bull* 2000, **26:** 351-366.

11. Committee for Medicinal Products for Human Use (CHMP). **Guideline on clinical investigation of medicinal products in the treatment of schizophrenia.** 2011. European Medicines Agency. 17-2-2011.

12. Crow TJ: **Molecular Pathology of schizophrenia: more than one disease process?** *Br Med J* 1980, **12:** 66-68.

13. Davis KL, Kahn RS, Ko G, Davidson M: **Dopamine in schizophrenia: a review and reconceptualization.** *Am J Psychiatry* 1991, **148:** 1474-1486.

14. Emsley RA, Roberts MC, Rataemane S, Pretorius J, Oosthuizen PP, Turner J *et al.*: **Ethnicity and treatment response in schizophrenia: a comparison of 3 ethnic groups.** *J Clin Psychiatry* 2002, **63:** 9-14.

15. Guy W. **Clinical Global Impression.** 1976. Rockville, MD, ECDEU Assessment Manual for Psychopharmacology, revised National Institute of Mental Health.

16. Kay SR: **Positive and negative syndromes in schizophrenia: assessment and research**. New York: BRUNNER/MAZEL INC.; 1991.

17. Kay SR, Fiszbein A, Opler LA: **The positive and negative syndrome scale (PANSS) for schizophrenia.** *Schizophr Bull* 1987, **13:** 261-276.

18. Kay SR, Opler LA, Lindenmayer JP: **The Positive and Negative Syndrome Scale (PANSS): rationale and standardization.** *British Journal of Psychiatry* 1989, **155(Suppl 7):** 59-65.

19. Kinon BJ, Chen L, scher-Svanum H, Stauffer VL, Kollack-Walker S, Zhou W *et al.*: **Early response to antipsychotic drug therapy as a clinical marker of subsequent response in the treatment of schizophrenia.** *Neuropsychopharmacology* 2010, **35:** 581-590.

20. Kolen MJ, Brennan RL: **Test Equating. Methods and Practices.** Springer, New York: 1995.

21. Kraepelin E: **Psychiatrie. Ein Lehrbuch für Studirende und Ärzte**, 6 edn. Barth, Leipzig: 1899.

22. 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, 7-14.

23. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR: **Linking the PANSS, BPRS, and CGI: clinical implications.** *Neuropsychopharmacology* 2006, **31:** 2318-2325.

24. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR: **What does the PANSS mean?** *Schizophr Res* 2005, **79:** 231-238.

25. Leucht S, Kissling W, Davis JM: **The PANSS Should Be Rescaled.** *Schizophr Bull* 2010, **36:** 461-462.

26. Levine SZ, Rabinowitz J: **Revisiting the 5 dimensions of the Positive and Negative Syndrome Scale.** *J Clin Psychopharmacol* 2007, **27:** 431-436.

27. 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.

28. Marder SR, Davis JM, Chouinard G: **The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials.** *J Clin Psychiatry* 1997, **58:** 538-546.

29. Meltzer HY, Bobo WV, Nuamah IF, Lane R, Hough D, Kramer M *et al.*: **Efficacy and tolerability of oral paliperidone extended-release tablets in the treatment of acute schizophrenia: pooled data from three 6-week, placebo-controlled studies.** *J Clin Psychiatry* 2008, **69:** 817-829.

30. Moher D, Liberati A, Tetzlaff J, Altman DG: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *BMJ* 2009, **339:b2535. doi: 10.1136/bmj.b2535.:** b2535.

31. Möller HJ, Laux G, Deister A: **Duale Reihe: Psychiatrie und Psychotherapie.** Thieme Verlag, Stuttgart: 2009.

32. Morrison AP, Hutton P, Wardle M, Spencer H, Barratt S, Brabban A *et al.*: **Cognitive therapy for people with a schizophrenia spectrum diagnosis not taking antipsychotic medication: an exploratory trial.** *Psychol Med* 2011, 1-8.

33. nil Yagcioglu AE, Kivircik Akdede BB, Turgut TI, Tumuklu M, Yazici MK, Alptekin K *et al.*: **A double-blind controlled study of adjunctive treatment with risperidone in schizophrenic patients partially responsive to clozapine: efficacy and safety.** *J Clin Psychiatry* 2005, **66:** 63-72.

34. Obermeier M, Mayr A, Schennach-Wolff R, Seemuller F, Moller HJ, Riedel M: **Should the PANSS Be Rescaled?** *Schizophr Bull* 2010, **36:** 455-460.

35. Obermeier M, Schennach-Wolff R, Meyer S, Moller HJ, Riedel M, Krause D *et al.*: **Is the PANSS used correctly? a systematic review.** *BMC Psychiatry* 2011, **11:113.:** 113.

36. Overall JE, Gorham DR: **The Brief Psychiatric Rating Scale.** *Psychological Reports* 1962, **10:** 799-812.

37. 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.

38. Schennach-Wolff R, Jager M, Mayr A, Meyer S, Kuhn KU, Klingberg S *et al.*: **Predictors of response and remission in the acute treatment of first-episode schizophrenia patients--is it all about early response?** *Eur Neuropsychopharmacol* 2011, **21:** 370-378.

39. Schennach-Wolff R, Obermeier M, Seemuller F, Jager M, Schmauss M, Laux G *et al.*: **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.

40. Schneider K: **Klinische Psychopathologie.** Thieme Verlag, Stuttgart: 1950.

41. Singh MM, Kay SR: **A comparative study of haloperidol and chlorpromazine in terms of clinical effects and therapeutic reversal with benztrompine in**

schizophrenia: Theoretical implications for potency differences among neuroleptics.** *Psychopharmacologia* 1975, **43:** 103-113.

42. Tran PV, Hamilton SH, Kuntz AJ, Potvin JH, Andersen SW, Beasley C, Jr. *et al*.: **Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders.** *J Clin Psychopharmacol* 1997, **17:** 407-418.

43. van NC, Schene AH, Koeter MW, Huxley PJ: **The Lancashire Quality of Life Profile: modification and psychometric evaluation.** *Soc Psychiatry Psychiatr Epidemiol* 2001, **36:** 36-44.

44. Wolwer W, Buchkremer G, Hafner H, Klosterkotter J, Maier W, Moller HJ *et al*.: **German research network on schizophrenia-bridging the gap between research and care.** *Eur Arch Psychiatry Clin Neurosci* 2003, **253:** 321-329.

First article:

Should the PANSS be rescaled?

Obermeier M, Mayr A, Schennach-Wolff R, Seemüller F, Möller HJ, Riedel M.

Should the PANSS Be Rescaled?

**Michael Obermeier[2], Andreas Mayr[1,2],
Rebecca Schennach-Wolff[2], Florian Seemüller[2],
Hans-Jürgen Möller[2], and Michael Riedel[2]**

[2]Department 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 ($PANSS_{99}$) with the corresponding baseline measurement ($PANSS_0$) can be analyzed and compared. Various effect measures have been discussed in statistical literature: Törnqvist et al[3] 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 Ayers[4] showed the high power of symmetrized percent change (PC) in statistical analyses. In the present article, we focus on the ordinary PC $100 \times (PANSS_{99} - PANSS_0)/PANSS_0$ because it is commonly used in schizophrenia research[5] 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] Peuskens[8]). 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 Adminstration[13]) 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 al[17]).

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

[1]To 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 ± 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 ($PANSS_0$) and end point ($PANSS_{99}$), 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 $PANSS_0$, 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 $PANSS_{99}$ on $PANSS_0$ for the real data set. The estimates of this model were used to generate $PANSS_{99}$ data for the 2 different subgroups on the basis of the simulated baseline data. As with this procedure, $PANSS_{99}$ and $PANSS_0$ would be perfectly correlated (cor = 1); additionally, a Gaussian noise (data from a normal distribution with $\mu = 0$ and a certain σ) was added on $PANSS_{99}$ to reach a correlation structure comparable to the real data. The greater the σ, the weaker is the correlation between $PANSS_{99}$ and $PANSS_0$ 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 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 ± 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 1. Real Data set; Group Effect Concerning PC?

	Mean PC Male (%)[a]	Mean PC Female (%)[a]	t Value[b]	P 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.
[a]Mean PC in male/female group from baseline to end point.
[b]Test statistic and P value (Wald test) of the estimated group effect (male/female) on PC in a linear model.

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

	20% Response		50% Response	
	z Value	P Value[a]	z Value[a]	P Value[a]
30 not subtracted	1.64	0.10	−0.05	0.96
30 subtracted	2.03	0.04	1.80	0.07

[a]Test statistic and P value (Wald test) of the estimated group effect (male/female) on response in a logistic model.

456
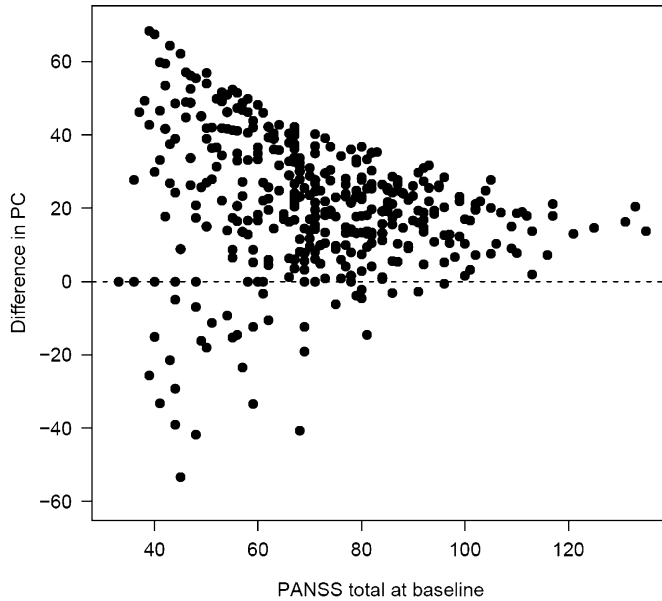
Fig. 1. Absolute Differences in Percent Changes Between Calculation Methods Depending on the Baseline Level in a Real Data Set. With decreasing baseline level, the differences between the calculation methods increase.

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 $PANSS_0$ baseline data, a right skewed (discretized) gamma distribution was most similar to the real data set. The relationship between $PANSS_0$ and $PANSS_{99}$

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 $PNASS_{99}$, had a σ of 15 and resulted in correlations between $PANSS_0$ and $PANSS_{99}$ from 0.39 to 0.59.

Table 3 shows some representative results regarding PC in the simulation study in relation to different levels of $PANSS_0$ 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.
[a]ID of simulation study.
[b]PANSS total: mean at baseline.
[c]Mean t value (Wald tests) of the estimated group effect on PC in a linear model.
[d]Empirical SD of differences in t values between both methods ($SD(t_1 - t_2)$).
[e]Number 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

	z Values[b]		Rates[c]		
ID[a]	30 Not Subtracted	30 Subtracted	30 Not Subtracted	30 Subtracted	Significant Change[d]
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

[a]ID of simulation study.
[b]Mean z value (Wald tests) of the estimated group effect on PC in a logistic model.
[c]Mean responder rates.
[d]Number 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: $|PC_{interval}| \leq |PC_{ratio}|$), 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 $PANSS_0$ influence the difference between results of both calculation methods: The higher the $PANSS_0$, the smaller is the slope of the nonlinear function mentioned. Therefore, with decreasing level of $PANSS_0$ 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

458

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.

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 Administzration. 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.

Second article:

Is the PANSS used correctly? a systematic review.

Obermeier M, Schennach-Wolff R, Meyer S, Möller HJ, Riedel M, Krause D, Seemüller F.

BMC Psychiatry

RESEARCH ARTICLE **Open Access**

Is the PANSS used correctly? a systematic review

Michael Obermeier[1*], Rebecca Schennach-Wolff[1], Sebastian Meyer[1], Hans-Jürgen Möller[1], Michael Riedel[1,2], Daniela Krause[1] and Florian Seemüller[1]

Abstract

Background: The PANSS (Positive and Negative Syndrome Scale) is one of the most important rating instruments for patients with schizophrenia. Nevertheless, there is a long and ongoing debate in the psychiatric community regarding its mathematical properties.

All 30 items range from 1 to 7 leading to a minimum total score of 30, implying that the PANSS is an interval scale. For such interval scales straightforward calculation of relative changes is not appropriate. To calculate outcome criteria based on a percent change as, e.g., the widely accepted response criterion, the scale has to be transformed into a ratio scale beforehand. Recent publications have already pointed out the pitfall that ignoring the scale level (interval vs. ratio scale) leads to a set of mathematical problems, potentially resulting in erroneous results concerning the efficacy of the treatment.

Methods: A Pubmed search based on the PRISMA statement of the highest-ranked psychiatric journals (search terms "PANSS" and "response") was carried out. All articles containing percent changes were included and methods of percent change calculation were analysed.

Results: This systematic literature research shows that the majority of authors (62%) actually appear to use incorrect calculations. In most instances the method of calculation was not described in the manuscript.

Conclusions: These alarming results underline the need for standardized procedures for PANSS calculations.

Keywords: PANSS, scale level, literature search

Background

The PANSS is currently the most established scale in patients with schizophrenia. For example in the high impact journal "Schizophrenia Bulletin" Kay's publication on the Positive and Negative Syndrome Scale (PANSS) for Schizophrenia is the most frequently cited article with more than 4000 citations (pubmed 05/2011) [1]. Despite its common use there still seems to be profound uncertainty within the psychiatric community regarding its mathematical properties. The pitfall relates to the calculation of proportions (including percent changes), which are used in common outcome criteria like response.

Dichotomized measures such as response can be understood more intuitively than mean values and are specifically endorsed by the European Medicines Agency

http://www.ema.europa.eu/htms/human/ich/ichefficacy.htm.

As pointed out in a previous paper [2], the PANSS is a 30 item interval scale ranging from 1-7 which implies that computations of ratios (e.g. percent changes, like calculation of XX% PANSS reduction from baseline to final endpoint) are not appropriate. Ignoring this fact leads to severe mathematical problems, resulting in an underestimation of the actual response rate and potentially even to erroneous results. Comparing results with and without PANSS scale level transformation into a ratio scale revealed that up to 50% of test decisions may differ [2]. In a comment on this article [3], Leucht et al. have cited such erroneous calculation methods as one reason for low response rates in studies on second generation antipsychotic drugs.

To avoid incorrect calculations the best solution would be to subtract the theoretical minimum (which is 30 for the total score), resulting in a score range starting from zero. Percent changes (PCs) have to be calculated

* Correspondence: Michael.Obermeier@med.uni-muenchen.de
[1]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University Munich, Nussbaumstrasse 7, 80336 Munich, Germany
Full list of author information is available at the end of the article

using this corrected version of the PANSS, which converts the PANSS into a ratio scale. Although Leucht et al. [4,5] have emphasized this necessity previously, the uncertainty in the psychiatric community remains.

In our previous report we already cited some articles performing the correction, as well as some others ignoring the pitfall. These examples also included approval studies of atypical antipsychotics, where a correct calculation would seem to be particularly important [6]. However, the mentioned articles were neither representative, nor did they give any answer to the scope of the problem. So far, knowledge concerning the relative frequency of incorrectly calculated PANSS PCs has been limited. If papers with erroneous calculations turn out to be negligible in comparison to similar publications as a whole, then most researchers seem to be aware of this pitfall. If not, we need to open a wider debate on this issue, because results of studies using different methods for the calculation of PCs can, strictly speaking, not be compared.

Thus, the aim of this review article is to further investigate the scope of incorrect PANSS calculations based on a systematic review of all articles published in the top ten journals with the highest impact factors in psychiatry, with a focus on the question: Is the PANSS used correctly?

Methods

All articles in this review were found by a systematic literature search in the top-ranked psychiatric journals using Pubmed http://www.pubmed.com based on the PRISMA statement [7]. The Impact Factor for psychiatric journals according to the 2008 Journal Citation Reports® Science Edition (Thomson Reuters, 2009) was used as ranking index. Journals focusing on topics not related to the PANSS and schizophrenia, such as Molecular Psychiatry or journals specialising in adolescent psychiatry, were excluded.

Based on these criteria, a predefined Pubmed search was carried out in the 10 highest-ranked journals entering the search terms "PANSS" and "response" with no restrictions regarding date of publication. The search term "response" was expected to be linked to the calculation of PCs in the PANSS.

Articles were included if they contained PCs in the PANSS in any form: Study inclusion criteria as well as outcome parameters were of interest, as well as continuous PCs and dichotomous response criteria. All articles containing PCs were included in this review and their methods of PC calculation were analyzed. The authors of articles with insufficient method descriptions were contacted (twice in case of no reply).

A classification was performed independently by two experienced researchers (MO and FS) into articles with PC as primary and those with PC as secondary outcome and into articles using PC as inclusion criteria. In case of disagreement a third researcher (SM) was consulted so that all articles could be satisfactorily classified.

Articles grouped according to their PC calculation method were sub-classified according to their year of publication, their outcome parameter and their particular citation number, using nonparametric, rank-based statistics and corresponding tests.

Results

The ten highest-ranked psychiatric journals according to their impact factor 2008 included three journals, which did not fit our search criteria (MOL PSYCHIATR, J CHILD PSYCHOL PSYC and J AM ACAD CHILD PSY). These three journals were therefore replaced by the three subsequent journals on the impact list (PSYCHOL MED, J PSYCHIATR RES, J NEUROL NEUROSUR). The search in Pubmed in January 2011 resulted in 68 publications including both terms, "PANSS" and "response". Of all articles, 39 actually used PANSS PC values ([8-46]) and for 33 articles the method of calculation could finally be determined. Table 1 shows the main results in detail.

In summary, in at least 62% of all publications (24 out of 39) the PANSS PC was calculated without the necessary score correction. The PC calculation method was rarely specified within the text. It was possible only in seven articles, to deduce the calculation method without correspondence with the authors: In two articles with score correction an explanation of the method was included and in five articles without correction the calculation method could be identified through an examination of the presented results.

Most of the articles were from the past few years (median:2007, range:1995-2010), without any noticeable difference (p = 0.23) between articles with (median:2008, range:1995-2010) and without score correction (median:2006.5, range:1998-2010). The number of citations ranged from 0 to 447 with a median of 18. As with the year of publication, there was no significant difference (p = 0.94) regarding the number of citations in the two groups. There is a significant negative rank correlation of -0.70 between citation number and publication year (p < 0.001).

Regarding the outcome classification of the articles, 33 of the 39 articles could be classified concordantly by researchers MO and FS, and in six cases a third researcher (SM) was consulted for the final decision. In twelve of the 39 publications the primary outcome was based on PC; in five (42%) of these corrected score values were used, five (42%) used uncorrected scores, and in two (17%) the method remained unclear. The majority of the articles found presented PCs as

Table 1 Summary of calculation methods in single journals

Abbreviated Journal Title (Impact Factor 2008)	No. of articles with correction	No. of articles without correction	No. of articles, unknown method
ARCH GEN PSYCHIAT (14.273)	**0**	**1**[8]	**1**[9]
AM J PSYCHIAT (10.545)	**1**[10]	**1**[11]	**0**
BIOL PSYCHIAT (8.672)	**1**[12]	**1**[13]	**0**
NEUROPSYCHOPHARMACOL (6.835)	**0**	**2**[14,15]	**1**[16]
SCHIZOPHRENIA BULL (6.592)	**1**[17]	**0**	**0**
BRIT J PSYCHIAT (5.077)	**3**[18-20]	**0**	**0**
J CLIN PSYCHIAT (5.053)	**3**[21-23]	**11**[24-34]	**1**[35]
PSYCHOL MED (4.718)	**0**	**1**[36]	**2**[37,38]
J PSYCHIATR RES (4.679)	**0**	**7**[39-45]	**1**[46]
J NEUROL NEUROSUR PS (4.622)	**0**	**0**	**0**

secondary outcomes:4 (15%) with correction, 19 (70%) without, and 4 (15%) articles with unknown status. There was no significant difference between outcome classification and method (p = 0.09).

Discussion

The influence of the PC calculation method on the results of double blind placebo controlled trials has already been described and quantified in detail in our previous article [2]. There are two main issues, which need to be considered: (1) Results of studies without correction cannot be compared to studies with correction. A 50% response criterion, for example, denotes two different facts: With corrected scores it corresponds to a 50% reduction of the measured symptoms, whereas without correction it corresponds to a 50% reduction of the score value, which is something very different. (2) Results are not only incomparable, but could even lead to different conclusions: While one method might reveal a significant treatment effect, the other might lead to the opposite result [2]. In articles with PC as primary outcome this is particularly problematic, since without correction even the main conclusion might be erroneous. A special issue in this context are approval studies, which are obliged to follow guidelines like the EMEA guidelines and therefore regularly include outcome measures with PCs. For one approval study [6] an erroneous calculation of the PANSS PC has already been shown [2].

In combination with the results of the present review it becomes even more apparent that there is a strong need for clarification in terms of the PANSS calculation: Although some authors use corrected scores, in the majority of cases the correction is not performed. Most importantly, the non-awareness of this problem is mirrored by the fact that only in two articles the score correction was described in the Methods section. This suggests that most researchers conducting schizophrenia trials are not even aware of this pitfall. Considering the fact that we probably did not identify all relevant articles in our literature search by focussing on the searching term of "response" one could assume that there are even more publications with incorrect PANSS calculations.

This is even more remarkable keeping in mind that the papers reviewed were published in high impact journals. So we can answer the question posed at the beginning of this article: Yes, the PANSS is used incorrectly!

What solutions can be made? First of all, it would be helpful to recalculate studies which have used the PANSS PC as primary outcome without correction.

For future work with the PANSS a consensus in the psychiatric research field is needed: Is it enough to correct the score every time PCs are used or should the PANSS be rescaled? Leucht et al., in their comment on our previous paper, prefer the radical solution: The PANSS items should be rescaled into a scale ranging from 0 to 6. This would be the most straightforward solution and could avoid future problems with PCs. Additionally, renaming the scale as e.g. "PANSS-0" or "PANSS (ratio version)", as suggested previously, could prevent new confusion, which might otherwise arise with different scale versions.

Conclusions

Again, we emphasize the necessity of further discussion and a broad consensus on future action in the psychiatric community. Until this is achieved we recommend that, for PANSS PC calculations, all researchers at least use the scale correction and include a short statement in the description of methods.

Author details
[1]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University Munich, Nussbaumstrasse 7, 80336 Munich, Germany. [2]Vinzenz von Paul Hospital, Psychiatry, Schwenninger Str. 55, 78628 Rottweil, Germany.

Authors' contributions
MO performed the analyses of the found articles and elaborated the conception of the manuscript, including a first draft. RS-W participated in the conception of the analysis and revised the manuscript critically. SM reviewed the included articles and assisted in the sequence alignment. H-JM, MR and DK revised the manuscript critically at each step of the analysis. FS reviewed the found articles and revised the manuscript critically. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Kay SR, Fiszbein A, Opler LA: **The positive and negative syndrome scale (PANSS) for schizophrenia.** *Schizophr Bull* 1987, **13**:261-276.
2. Obermeier M, Mayr A, Schennach-Wolff R, Seemuller F, Moller HJ, Riedel M: **Should the PANSS Be Rescaled?** *Schizophr Bull* 2010, **36**:455-460.
3. Leucht S, Kissling W, Davis JM: **The PANSS Should Be Rescaled.** *Schizophr Bull* 2010, **36**:461-462.
4. 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.
5. 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, 7-14.
6. FDA: **Application No.:021253. FDA.** *FDA* 2004 [http://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/21253_Zyprexa.TOC.cfm], Ref Type: Electronic Citation.
7. Moher D, Liberati A, Tetzlaff J, Altman DG: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *BMJ* 2009, **339**:b2535.
8. Heresco-Levy U, Javitt DC, Ermilov M, Mordel C, Silipo G, Lichtenstein M: **Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia.** *Arch Gen Psychiatry* 1999, **56**:29-36.
9. Breier A, Meehan K, Birkett M, David S, Ferchland I, Sutton V, *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.
10. Emsley R, Rabinowitz J, Medori R: **Time course for antipsychotic treatment response in first-episode schizophrenia.** *Am J Psychiatry* 2006, **163**:743-745.
11. Sikich L, Frazier JA, McClellan J, Findling RL, Vitiello B, Ritz L, *et al*: **Double-blind comparison of first- and second-generation antipsychotics in early-onset schizophrenia and schizo-affective disorder: findings from the treatment of early-onset schizophrenia spectrum disorders (TEOSS) study.** *Am J Psychiatry* 2008, **165**:1420-1431.
12. Marder SR, Kramer M, Ford L, Eerdekens E, Lim P, Eerdekens M, *et al*: **Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study.** *Biol Psychiatry* 2007, **62**:1363-1370.
13. Campbell DB, Ebert PJ, Skelly T, Stroup TS, Lieberman J, Levitt P, *et al*: **Ethnic stratification of the association of RGS4 variants with antipsychotic treatment response in schizophrenia.** *Biol Psychiatry* 2008, **63**:32-41.
14. Kegeles LS, Slifstein M, Frankle WG, Xu X, Hackett E, Bae SA, *et al*: **Dose-occupancy study of striatal and extrastriatal dopamine D2 receptors by aripiprazole in schizophrenia with PET and [18F]fallypride.** *Neuropsychopharmacology* 2008, **33**:3111-3125.
15. Sechter D, Peuskens J, Fleurot O, Rein W, Lecrubier Y: **Amisulpride vs. risperidone in chronic schizophrenia: results of a 6-month double-blind study.** *Neuropsychopharmacology* 2002, **27**:1071-1081.
16. Kinon BJ, Chen L, scher-Svanum H, Stauffer VL, Kollack-Walker S, Zhou W, *et al*: **Early response to antipsychotic drug therapy as a clinical marker of

subsequent response in the treatment of schizophrenia.** *Neuropsychopharmacology* 2010, **35**:581-590.
17. Levine SZ, Rabinowitz J: **Trajectories and Antecedents of Treatment Response Over Time in Early-Episode Psychosis.** *Schizophr Bull* 2008.
18. Haas M, Eerdekens M, Kushner S, Singer J, Augustyns J, Quiroz J, *et al*: **Efficacy, safety and tolerability of two dosing regimens in adolescent schizophrenia: double-blind study.** *Br J Psychiatry* 2009, **194**:158-164.
19. 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.
20. Schennach-Wolff R, Seemuller FH, Mayr A, Maier W, Klingberg S, Heuser I, *et al*: **An early improvement threshold to predict response and remission in first-episode schizophrenia.** *Br J Psychiatry* 2010, **196**:460-466.
21. Fleischhacker WW, Eerdekens M, Karcher K, Remington G, Llorca PM, Chrzanowski W, *et al*: **Treatment of schizophrenia with long-acting injectable risperidone: a 12-month open-label trial of the first long-acting second-generation antipsychotic.** *J Clin Psychiatry* 2003, **64**:1250-1257.
22. Volpi S, Potkin SG, Malhotra AK, Licamele L, Lavedan C: **Applicability of a genetic signature for enhanced iloperidone efficacy in the treatment of schizophrenia.** *J Clin Psychiatry* 2009, **70**:801-809.
23. Canuso CM, Lindenmayer JP, Kosik-Gonzalez C, Turkoz I, Carothers J, Bossie CA, *et al*: **A randomized, double-blind, placebo-controlled study of 2 dose ranges of paliperidone extended-release in the treatment of subjects with schizoaffective disorder.** *J Clin Psychiatry* 2010, **71**:587-598.
24. Meltzer HY, Bobo WV, Nuamah IF, Lane R, Hough D, Kramer M, *et al*: **Efficacy and tolerability of oral paliperidone extended-release tablets in the treatment of acute schizophrenia: pooled data from three 6-week, placebo-controlled studies.** *J Clin Psychiatry* 2008, **69**:817-829.
25. Glick ID, Poyurovsky M, Ivanova O, Koran LM: **Aripiprazole in schizophrenia patients with comorbid obsessive-compulsive symptoms: an open-label study of 15 patients.** *J Clin Psychiatry* 2008, **69**:1856-1859.
26. Kahn RS, Schulz SC, Palazov VD, Reyes EB, Brecher M, Svensson O, *et al*: **Efficacy and tolerability of once-daily extended release quetiapine fumarate in acute schizophrenia: a randomized, double-blind, placebo-controlled study.** *J Clin Psychiatry* 2007, **68**:832-842.
27. Chan HY, Lin WW, Lin SK, Hwang TJ, Su TP, Chiang SC, *et al*: **Efficacy and safety of aripiprazole in the acute treatment of schizophrenia in Chinese patients with risperidone as an active control: a randomized trial.** *J Clin Psychiatry* 2007, **68**:29-36.
28. Zhong KX, Sweitzer DE, Hamer RM, Lieberman JA: **Comparison of quetiapine and risperidone in the treatment of schizophrenia: A randomized, double-blind, flexible-dose, 8-week study.** *J Clin Psychiatry* 2006, **67**:1093-1103.
29. nil Yagcioglu AE, Kivircik Akdede BB, Turgut TI, Tumuklu M, Yazici MK, Alptekin K, *et al*: **A double-blind controlled study of adjunctive treatment with risperidone in schizophrenic patients partially responsive to clozapine: efficacy and safety.** *J Clin Psychiatry* 2005, **66**:63-72.
30. Lane HY, Chang YC, Cheng YC, Liu GC, Lin XR, Chang WH: **Effects of patient demographics, risperidone dosage, and clinical outcome on body weight in acutely exacerbated schizophrenia.** *J Clin Psychiatry* 2003, **64**:316-320.
31. Lindenmayer JP, Czobor P, Volavka J, Lieberman JA, Citrome L, Sheitman B, *et al*: **Olanzapine in refractory schizophrenia after failure of typical or atypical antipsychotic treatment: an open-label switch study.** *J Clin Psychiatry* 2002, **63**:931-935.
32. Emsley RA, Roberts MC, Rataemane S, Pretorius J, Oosthuizen PP, Turner J, *et al*: **Ethnicity and treatment response in schizophrenia: a comparison of 3 ethnic groups.** *J Clin Psychiatry* 2002, **63**:9-14.
33. Dossenbach MR, Kratky P, Schneidman M, Grundy SL, Metcalfe S, Tollefson GD, *et al*: **Evidence for the effectiveness of olanzapine among patients nonresponsive and/or intolerant to risperidone.** *J Clin Psychiatry* 2001, **62(Suppl 2)**:28-34.
34. Lindenmayer JP, Iskander A, Park M, Apergi FS, Czobor P, Smith R, *et al*: **Clinical and neurocognitive effects of clozapine and risperidone in treatment-refractory schizophrenic patients: a prospective study.** *J Clin Psychiatry* 1998, **59**:521-527.
35. Koenigsberg HW, Reynolds D, Goodman M, New AS, Mitropoulou V, Trestman RL, *et al*: **Risperidone in the treatment of schizotypal personality disorder.** *J Clin Psychiatry* 2003, **64**:628-634.

36. Case M, Stauffer VL, scher-Svanum H, Conley R, Kapur S, Kane JM, *et al*: **The heterogeneity of antipsychotic response in the treatment of schizophrenia.** *Psychol Med* 2010, 1-10.
37. Mizrahi R, Korostil M, Starkstein SE, Zipursky RB, Kapur S: **The effect of antipsychotic treatment on Theory of Mind.** *Psychol Med* 2007, **37**:595-601.
38. Copolov DL, Link CG, Kowalcyk B: **A multicentre, double-blind, randomized comparison of quetiapine (ICI 204,636, 'Seroquel') and haloperidol in schizophrenia.** *Psychol Med* 2000, **30**:95-105.
39. Musil R, Spellmann I, Riedel M, Dehning S, Douhet A, Maino K, *et al*: **SNAP-25 gene polymorphisms and weight gain in schizophrenic patients.** *J Psychiatr Res* 2008, **42**:963-970.
40. Lipkovich I, Deberdt W, Csernansky JG, Buckley P, Peuskens J, Kollack-Walker S, *et al*: **Predictors of risk for relapse in patients with schizophrenia or schizoaffective disorder during olanzapine drug therapy.** *J Psychiatr Res* 2007, **41**:305-310.
41. Bozina N, Kuzman MR, Medved V, Jovanovic N, Sertic J, Hotujac L: **Associations between MDR1 gene polymorphisms and schizophrenia and therapeutic response to olanzapine in female schizophrenic patients.** *J Psychiatr Res* 2008, **42**:89-97.
42. Prikryl R, Ceskova E, Kasparek T, Kucerova H: **Neurological soft signs, clinical symptoms and treatment reactivity in patients suffering from first episode schizophrenia.** *J Psychiatr Res* 2006, **40**:141-146.
43. Peet M, Horrobin DF: **A dose-ranging exploratory study of the effects of ethyl-eicosapentaenoate in patients with persistent schizophrenic symptoms.** *J Psychiatr Res* 2002, **36**:7-18.
44. Girgis RR, Merrill DB, Vorel SR, Edward K, Kimberly P, You M, *et al*: **Aripiprazole versus haloperidol treatment in early-stage schizophrenia.** *J Psychiatr Res* 2010.
45. Giegling I, Drago A, Schafer M, Moller HJ, Rujescu D, Serretti A: **Interaction of haloperidol plasma level and antipsychotic effect in early phases of acute psychosis treatment.** *J Psychiatr Res* 2010, **44**:487-492.
46. McEvoy JP, Daniel DG, Carson WH, McQuade RD, Marcus RN: **A randomized, double-blind, placebo-controlled, study of the efficacy and safety of aripiprazole 10, 15 or 20 mg/day for the treatment of patients with acute exacerbations of schizophrenia.** *J Psychiatr Res* 2007, **41**:895-905.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-244X/11/113/prepub

Third article:

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.

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, Rüther 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

Rebecca Schennach-Wolff, MD,* Michael Obermeier,* Florian Seemüller, MD,*
Markus Jäger, MD,* Max Schmauss,† Gerd Laux,‡ Herbert Pfeiffer,§ Dieter Naber,||
Lutz G. Schmidt,¶ Wolfgang Gaebel,# Joachim Klosterkötter,** Isabella Heuser,†† Wolfgang Maier,‡‡
Matthias R. Lemke,§§ Eckart Rüther,|||| Stefan Klingberg,¶¶ Markus Gastpar,## Rolf R. Engel,*
Hans-Jürgen Möller,* and Michael Riedel****

Background: Linking of the Clinical Global Impression (CGI) Scale and the Positive and Negative Syndrome Scale (PANSS) was performed within a naturalistic sample. Furthermore, these linking results were compared with those derived from randomized controlled trials to examine if the baseline severity might influence the linking results.

Methods: Biweekly PANSS and CGI ratings were performed from admission to discharge in 398 schizophrenia patients treated within a naturalistic study. Equipercentile linking was performed using the statistical program, R 2.8.1. To evaluate how the naturalistic study design would influence linkage results, a so-called study sample was computed with patients of the naturalistic study fulfilling common inclusion criteria of randomized controlled trials (n = 199). Patients not fulfilling these criteria (less ill sample) and those fulfilling the criteria (study sample) were compared using confidence intervals.

Results: We found a considerable difference between the linking of the CGI severity score and the PANSS total score comparing the less ill sample and the study sample. Being considered "mildly ill" at admission in the less ill sample corresponded to a PANSS total score of 47 points and to a PANSS total score of 67 points in the study sample. Considering the linking of the CGI improvement score and PANSS changes, similar results were found for CGI improvement ratings ranging from "very much improved" to "minimally improved."

Conclusions: Despite considerable differences, a 50% PANSS reduction was found to correspond to a clinical rating of much improved, which seems to be a suitable definition for response in clinical drug trials.

Today, the clinical implications of the Positive and Negative Syndrome Scale (PANSS),[1] one of the most frequently used rating scales to assess schizophrenic psychopathology, are not yet fully understood. To improve the understanding of PANSS scores and their percentage improvement from a clinical point of view, corresponding points for simultaneous ratings of the PANSS and the Clinical Global Impression Scale (CGI) were analyzed. The CGI is thought to be understood more intuitively describing the patient's overall clinical state helping to increase the understanding of specific PANSS ratings.

However, to this day, predominantly, patients in clinical pharmaceutical trials requiring inclusion criteria with a predefined minimum symptom severity have been analyzed.[1–3] It might be hypothesized that this selection bias might increase corresponding values limiting generalizability with a real-world situation, which includes patients of all severity grades. It is, thus, unclear if currently proposed results differ when examining different patient samples. If PANSS and CGI score would correspond differently depending on the patient sample, this might have consequences for clinical practice.

Therefore, the aim of the present analysis was to link the PANSS and the CGI within a naturalistic treatment setting with broad inclusion and exclusion criteria. In addition, to test whether the patient sample itself would have an impact on the linking, these results were compared with an artificial study sample.

MATERIALS AND METHODS

Subjects and Assessments

Data were collected in a multicenter follow-up program (German Research Network on Schizophrenia).[4] All patients with the diagnosis of schizophrenia, schizophreniform disorder, delusional disorder, and schizoaffective disorder according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*; American Psychiatric Association, 1994) and aged between 18 and 65 years were selected for inclusion. The exclusion criteria were a head injury, a history of major medical illness, and alcohol or drug dependency. An informed written consent had to be provided. The study protocol was approved by the local ethics committee. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia[5] and the Clinical Global Impression Severity Scale (CGI-S)[6] were
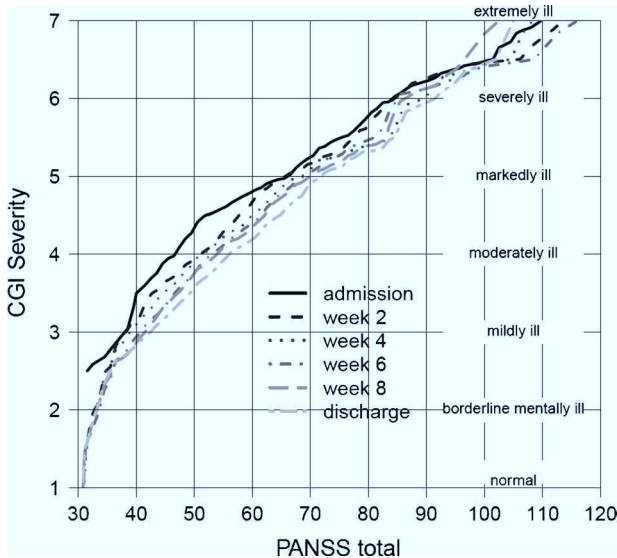
35

FIGURE 1. Linking of CGI-S score and PANSS total score (whole patient sample).

applied. The Brief Psychiatric Rating Scale (BPRS) was extracted from the PANSS to form the study sample. All raters had been trained using the applied scales. A high inter-rater reliability was achieved (analysis of variance, intraclass correlation coefficient > 0.8).

Statistical Analysis

To examine the degree of correlation to perform the equipercentile linking, the correlations between the CGI and PANSS were analyzed. To correlate the CGI-S rating and the PANSS total score, the Spearman correlation coefficient was applied, and for correlations between the CGI improvement (CGI-I) score and changes in the PANSS total score, the Pearson correlation coefficient was applied.

For the present analysis, equipercentile linking was used, a technique that identifies those scores on both measures that have

the same percentile rank. First, percentile rank functions are calculated for both measures. Using the percentile rank function of one variable and the inverse percentile function of the other, one finds a score on the one variable that has the same percentile rank for every score of the other variable, respectively. The exact statistical formula has been described by Kolen and Brennan.[7]

The linking was performed for the severity scores of the two scales and for the improvements for each study visit from baseline biweekly until discharge. To set the improvement or worsening during the course of treatment in relation to the psychopathological severity at admission in addition to absolute changes, percentage changes are also shown. To evaluate if the study sample itself might influence the linkage results, an artificial study sample was created out of the naturalistic sample. To do this, the commonly used inclusion criteria of randomized controlled trials (RCTs)[8–10] were applied to the whole patient sample. Patients scoring 4 or higher on the CGI-S score and concurrently 42 points or more on the BPRS were included in the so-called study sample, and patients not fulfilling these criteria were grouped into a "less ill sample." The linking of the less ill sample and the study sample was compared by confidence intervals using classical nonparametric bootstrap. All statistical analyses were performed using the statistical program, R2.8 .1.[11]

RESULTS

Patients

In the entire multicenter study, 474 patients were enrolled. 46 patients had to drop out for different reasons, another 28 patients were excluded because they were discharged from the hospital within 7 days after admission, and 2 patients were also excluded because of missing CGI and PANSS values. The remaining 398 patients participated in the study. The created study sample consisted of 199 patients achieving a BPRS score of 42 points or higher and concurrently a CGI-S score of 4 or higher. This study sample featured significantly more male patients ($P = 0.04$), worsened significantly less often during inpatient treatment ($P = 0.01$) and was significantly less often treated with an atypical antipsychotic monotherapy ($P = 0.04$)

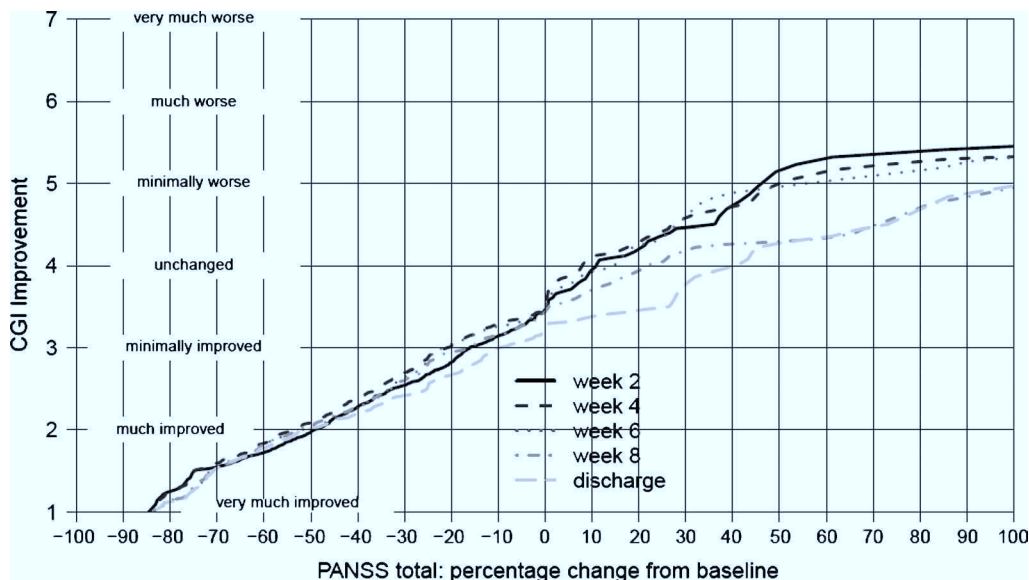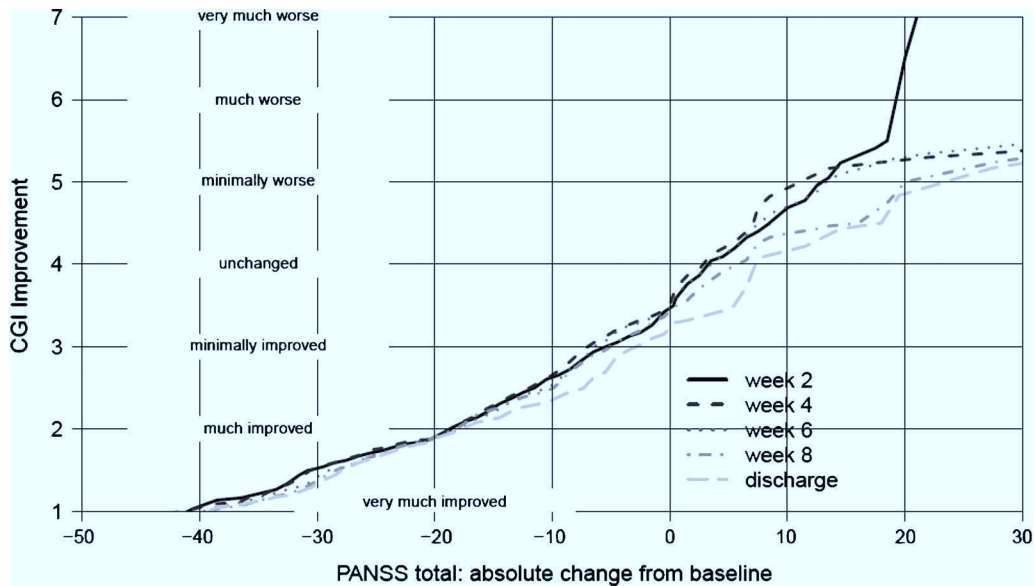**FIGURE 2.** Linking of CGI-I and PANSS percentage changes (whole patient sample).

FIGURE 3. Linking of CGI-I and PANSS absolute changes (whole patient sample).

compared with patients in the less ill sample also comprising 199 patients.

ASSESSMENTS

PANSS and CGI Correlations

Correlation coefficients between the PANSS total score and the CGI-S rating were 0.41 for admission, 0.62 for week 2, 0.63 for week 4, 0.62 for week 6, 0.62 for week 8, and 0.60 for discharge. Correlation coefficients between PANSS changes and the CGI-I score were 0.58 for week 2, 0.64 for week 4, 0.61 for week 6, 0.59 for week 8, and 0.56 for discharge.

Linking the PANSS Total Score and CGI-S Within a Naturalistic Sample

Results of the linking between the PANSS total score and the CGI-S of the less ill sample at admission, weeks 2 to 8, and discharge are shown in Figure 1. "Moderately ill" on the CGI-S (= 4) corresponded to a PANSS total score of 47 points at baseline; 51 points at week 2; 52 points at week 4; 53 points at week 6; 54 points at week 8; and 56 points at discharge. Being considered "markedly ill" on the CGI-S (= 5) corresponded to a PANSS total score of 66 points at admission and week 2; 69 points at weeks 4 and 6; 70 points at week 8; and 72 points for discharge. A score of "severely ill" on the CGI-S (= 6) corresponded to a PANSS total score of 84 points at admission; 84.75 points at week 2; 90.50 points at week 4; 85 points at week 6; 87 points at week 8; and 90.25 points at discharge.

Linking the Percentage Change of the PANSS and CGI-I Within a Naturalistic Sample

The linking between the CGI-I and the percentage PANSS change from admission to week 2 up to week 8 and discharge is displayed in Figure 2. Ratings of "very much improved" (CGI-I, 1) corresponded to a percentage PANSS reduction of 84% at weeks 2 and 4 and of 83% at weeks 6 and 8 and at discharge. Ratings of "much improved" (CGI-I, 2) corresponded to a percentage reduction of 49% at week 2; 54% at week 4; 52% at week 6; 53% at week 8; and 2% at discharge. The rating of "unchanged" of the CGI-I scale corresponded to a percentage increase (equal to worsening) of the PANSS score.

Linking the Absolute Change of the PANSS and CGI-I Within a Naturalistic Sample

The linking of the CGI-I and the absolute change of the PANSS is shown in Figure 3. Ratings of very much improved (CGI-I, 1) corresponded to an absolute PANSS change of 41.25 points at week 2 and 41 points at week 4; 40 points at week 6; and 39 points at week 8 and at discharge. Ratings of much improved (CGI-I, 2) corresponded to an absolute change of 19 points at weeks 2 and 4 and 18 points at weeks 6 and 8 and at discharge.

Confidence Intervals to Compare Linking Results of the Less Ill Sample and Study Sample

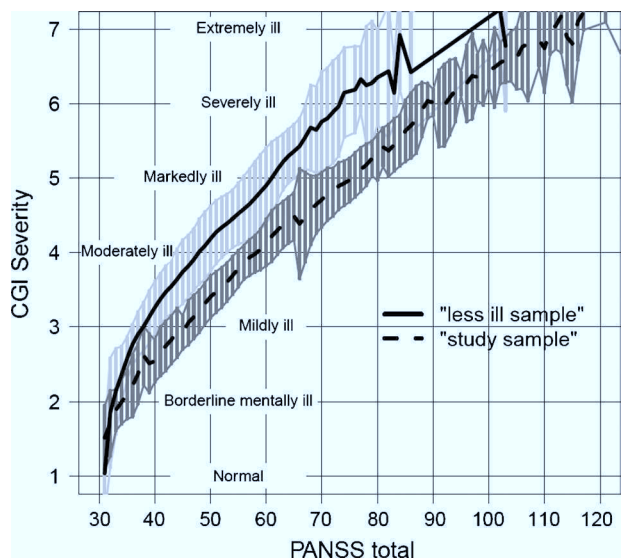- Severity scores (linkage between CGI-S and PANSS total score)

FIGURE 4. Confidence intervals comparing the linkage of the CGI-S score and PANSS total score between the less ill sample and the study sample.
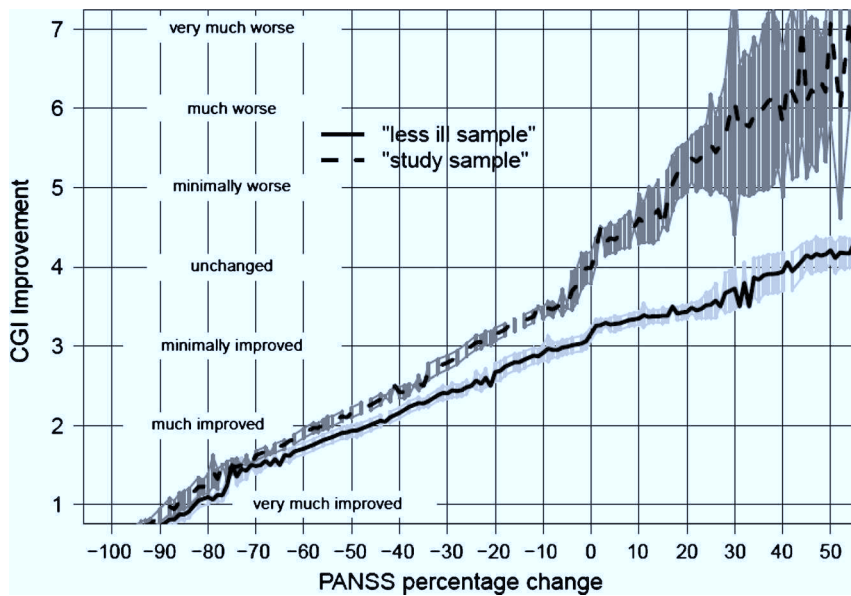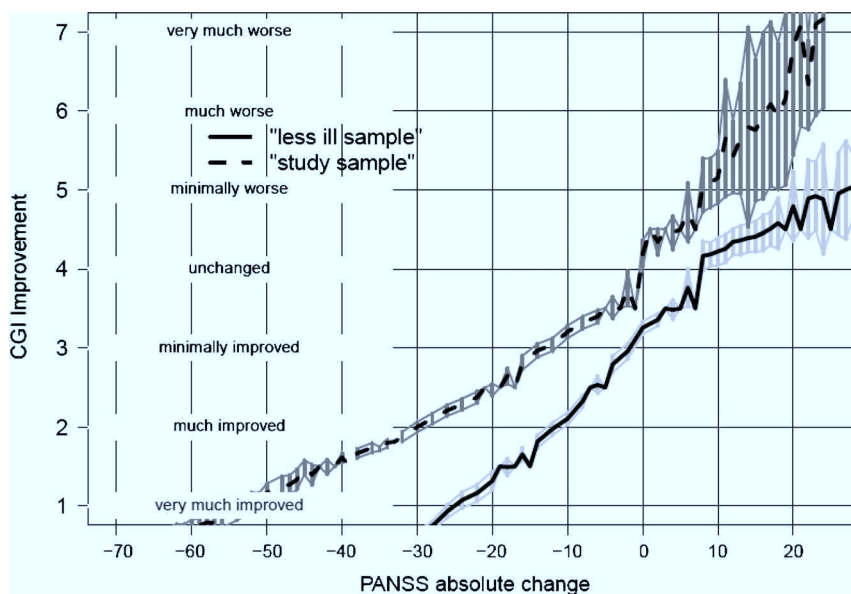
37

FIGURE 5. Confidence intervals comparing the linkage of the CGI-I score and PANSS percentage change between the less ill sample and the study sample.

Both confidence intervals suggest a satisfying goodness of estimation by featuring rather small distributions (Fig. 4). The confidence intervals of the 2 different patient samples overlap not until a CGI-I score of 6. This indicates that the choice of the sample has a substantial influence of the results of the linking.
• Improvement scores (linkage between CGI-I and percentage PANSS change)

Confidence intervals were furthermore computed for both patient samples regarding the improvement scores (Fig. 5). In the less ill sample, the CGI-I rating of unchanged corresponded to a percentage worsening of the PANSS at every assessed time point, whereas in the study sample, the CGI-I rating of unchanged corresponded to a percentage improvement until

week 8. Starting with the rating of unchanged, the confidence interval dispersed considerably, mirroring that there were considerably fewer patients with a worsening of symptoms in the study sample compared with the less ill sample, which explains the stretched confidence interval of the study sample regarding the worsening of symptoms.
• Improvement scores (linkage between CGI-I and absolute PANSS change)

Given that the patients in the study sample featured a higher PANSS total score at admission, the absolute change of patients that very much or much improved was found to be considerably higher than in the less ill sample mirrored in the dispersed confidence intervals (Fig. 6). Consistent with the linking of

FIGURE 6. Confidence intervals comparing the linkage of the CGI-I score and PANSS absolute change between the less ill sample and the study sample.

the percentage improvement diverged confidence intervals were also found starting with ratings of an unchanged psychopathological condition.

DISCUSSION

Linking of a Naturalistic Study Sample

This is the first analysis presenting linking results of the CGI and the PANSS derived from a naturalistic study reviling surprisingly rather low PANSS scores to a corresponding CGI when linking the CGI-S and the PANSS total score. For example, a CGI-S rating of moderately ill corresponded to a PANSS total score of 47 points at admission, increasing up to 56 points at discharge. Because the corresponding PANSS total score increases with the study duration, this might suggest that the study registrars rated more stringent with the course of the study being possibly more tolerant regarding symptom severity at admission. Another explanation might be that at baseline, physicians tended to rate the CGI more stringently, allowing more symptoms at discharge despite rating the patient to have improved.

When comparing these linkage results to other reports, our PANSS ratings are almost 30 to 40 points lower in relation to the corresponding CGI rating.[1,3] In our own analysis, a baseline CGI-S score of moderately ill referred to a PANSS total score of 47 points, whereas in the PANSS and CGI linkage study by, for example, Leucht et al,[1] a CGI-S score of moderately ill corresponded to a PANSS total score of 78 points at baseline. One explanation for these differing results might lie in the patient populations analyzed. In the study by Leucht et al, the data examined included mainly studies with a symptom-severity inclusion criterion, therefore analyzing at least moderately ill patients at study entry[1] resulting in a mean PANSS total score of 94 ± 19 points. In the present naturalistic study, the PANSS mean score at admission was 71 ± 19 points, and the widely differing mean scores already suggest different linkage results.

Comparing Linking Results of a Naturalistic Sample to Those of RCTs

By analyzing the confidence intervals, we wanted to demonstrate the difference between the corresponding linkage results of a naturalistic sample and a study sample derived from RCTs. The confidence interval for the CGI-S score and the PANSS total score comparing both study populations shows that the corresponding ratings vary the least regarding severely ill patients, as the confidence interval covers starting with a CGI rating of severely ill. This suggests that there is an understanding between physicians of what a severely ill patient is compared to a patient with less prominent and less interfering symptoms.

Comparing the linking of the improvement and worsening scores of our own 2 samples, we found a considerable difference in CGI-I rating of "unchanged." For example, a CGI rating of "unchanged" at week 2 corresponded to a worsening in the PANSS total score by an increase of 11% in the naturalistic analysis, but the same CGI rating corresponded to a PANSS total score improvement by 1 percentage point in the study sample. Not until a patient was rated as "minimally worse" did the PANSS change correspond to a percentage increase in the study sample.

The results of the study sample are in line with other literature results of controlled studies[1,3] indicating that in rating acutely symptomatic patients, the rater might expect a relatively greater improvement to be clinically meaningful. On the other hand, when a patient has more chronic symptoms, as is often the case in naturalistic studies, the rater might tolerate a worsening

of symptoms on the PANSS without evaluating this as clinically meaningful.

However, because there are still differences between comparative literature and our own study sample, there seem to be more influencing factors on linking results as only the baseline inclusion criteria. One gets the impression that present results of our less ill sample and our study sample vary by almost 1 point on the CGI-S score in corresponding PANSS ratings compared to other link analyses. One explanation might be rater specific. Present data were assessed in Germany, whereas comparative studies were often performed in the United States or in other non-European countries. The association of the cultural background and rating behavior is a well-known phenomenon in psychiatry.[12]

In addition, the statistical procedure of linking should be kept in mind when discussing the current inconsistency of data. Because in the link analysis, the PANSS and CGI scores are examined in a pooled way, meaning that the patient's individual corresponding PANSS and CGI rating is broken up, which seems to be problematic in some cases. For example, a patient is hearing one voice occasionally telling him to kill himself. This patient would not score very high on the PANSS scale because he only hears one voice occasionally; however, clinically, he would probably be rated severely ill, as he is in great danger regarding suicidal actions. On the other hand, patients with multiple symptoms scoring high on the PANSS scale might be able to cope well and continue everyday life being rated only mildly ill on the CGI scale. In these cases, PANSS and CGI rating diverge immensely, but in the individual patient, they are clinically reasonable and meaningful. Taking away the individual connection of these two ratings as performed in linking might result in two very dispersing ratings.

Clinical Implications

As demonstrated, different patient populations obviously lead to different linking results mainly based on different PANSS and CGI scores at baseline. This result is not very surprising and has already been expected by Leucht et al[1] in one of the first linking analyses. However, we believe that this is an important result limiting generalizability of previous linking studies, which should be kept in mind when interpreting the results of corresponding PANSS and CGI data.

Besides, the varying ratings suggest that there is no international consensus of PANSS and CGI ratings. In the less ill sample, the CGI was rated rather strictly in relation to fewer symptoms on the PANSS scale. Comparative literature data report this almost the other way around, as several studies found high PANSS scores with rather moderate CGI scores. However, maybe an equal and stabile correspondence between these 2 scales is not possible and possibly not necessary. Because the aforementioned clinical examples on PANSS and CGI ratings of the patient hearing one voice occasionally telling him to kill himself or the patient with multiple symptoms underline that the scales are not synonymous and the correlations are not very high.[13] Therefore, when performing clinical trials, both rating scales should concurrently be implemented.

In agreement with previous results, we also found a 49% to 50% PANSS total score reduction to correspond to a clinical status of much improved, which has been found to adequately mirror response in clinical trials.[1,3] This underlines the importance of implementing a 50% PANSS total score reduction as standard response criterion in clinical schizophrenia trials finding it to correspond to a clinical measure of much improved independent of the analyzed study sample.

Strengths and Limitations

One strength of this study is that it is the first to present linking results within a naturalistic setting. However, this is only a small patient sample derived from one study. In addition, because of the multicenter design, different raters were involved in the clinical evaluation. Because equipercentile linking assumes the measurement of the same underlying global trait (ie, global symptoms), it does not permit an examination of global (CGI) to individual (PANSS) symptoms.[14] Future studies based on a large database of patients with all degrees of severity in a sufficient number including the extremes are warranted to contribute further to the understanding of the PANSS and CGI and their clinical implications.

REFERENCES

1. Leucht S, Kane JM, Kissling W, et al. What does the PANSS mean? *Schizophr Res*. 2005;79:231–238.

2. Leucht S, Kane JM, Etschel E, et al. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology*. 2006;31: 2318–2325.

3. Levine SZ, Rabinowitz J, Engel R, et al. Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr Res*. 2008;98:318–322.

4. Wolwer W, Buchkremer G, Hafner H, et al. German research network on schizophrenia—bridging the gap between research and care. *Eur Arch Psychiatry Clin Neurosci*. 2003;253:321–329.

5. Kay SR, Opler LA, Lindenmayer JP. Reliability and validity of the positive and negative syndrome scale for schizophrenics. *Psychiatry Res*. 1988;23:99–110.

6. Guy W. Clinical Global Impressions. *ECDEU Assessment Manual for Psychopharmacology* 1976; National Institute of Mental Health, Rockville, MD; revised.

7. Kolen MJ, Brennan RL. *Test Equating: Methods and Practices*. New York, NY: Springer; 1995.

8. Breier A, Berg PH, Thakore JH, et al. Olanzapine versus ziprasidone: results of a 28-week double-blind study in patients with schizophrenia. *Am J Psychiatry*. 2005;162:1879–1887.

9. Beasley CM Jr, Sanger T, Satterlee W, et al. Olanzapine versus placebo: results of a double-blind, fixed-dose olanzapine trial. *Psychopharmacology (Berl)*. 1996;124:159–167.

10. Tran PV, Hamilton SH, Kuntz AJ, et al. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. *J Clin Psychopharmacol*. 1997;17:407–418.

11. R Development Core Team. *A Language and Environment for Statistical Computing*. URL: http/:www.R-project.org. Vienna, Austria: Foundation for Statistical Computing; 2008.

12. Mackin P, Targum SD, Kalali A, et al. Culture and assessment of manic symptoms. *Br J Psychiatry*. 2006;189:379–380.

13. Rabinowitz J, Mehnert A, Eerdekens M. To what extent do the PANSS and CGI-S overlap? *J Clin Psychopharmacol*. 2006;26:303–307.

14. Bernstein I.H., Teng G. Factoring items and factoring scales are different: spurious evidence for multidimensionality due to item categorization. *Psychol Bull*. 1989;105(3):467–477.

Danksagung

Diese Doktorarbeit entstand in den Jahren 2009 bis 2011 an der Psychiatrischen Klinik der Ludwig-Maximilians Universität München. Ich möchte die Gelegenheit nutzen, den Personen zu danken, durch deren direkte oder indirekte Unterstützung ein gutes Gelingen erst möglich wurde.

Ein großer Dank geht an Herrn Prof. Dr. Hans-Jürgen Möller, der sowohl als Doktorvater als auch als Ideengeber und konstruktiv-kritischer Begleiter aller drei Artikel einen nicht zu überschätzenden Anteil an dieser Arbeit hat. Die stets vertrauensvolle Zusammenarbeit bildete die Grundlage für meine Forschungen und ermöglichte somit erst diese Arbeit.

Ebenfalls bedanken möchte ich mich bei Herrn Prof. Dr. Rolf Engel, der als Betreuer der Dissertation stets ein sehr hilfreicher Ansprechpartner und mit seinem guten Rat maßgeblich am Erfolg beteiligt war.

Für die gute Zusammenarbeit generell möchte ich mich bei PD Dr. Michael Riedel bedanken: sein Vertrauen in meine Arbeit und seine stets hilfreiche Unterstützung haben die angenehme und produktive Atmosphäre entstehen lassen, die für gute wissenschaftliche Arbeit unerlässlich ist.

Ein besonderer Dank geht an meine Mitstreiter in der Katamnese-Forschungsgruppe: wissenschaftliche Diskussionen mit Dr. Rebecca Schennach, Dr. Florian Seemüller und Sebastian Meyer erwiesen sich stets als Genuss und überdies als sehr ertragreich und zielführend. Ihre kritische und fundierte Expertise zu meinen Arbeiten war essentiell für den Fortschritt in meiner Forschungsarbeit. Rebecca Schennach gebührt obendrein ein zusätzlicher Dank in ihrer Eigenschaft als Erstautorin eines in diese Dissertation eingebundenen Artikels.

Für die Ko-Autorenschaft generell, für das Sammeln der Daten und für die Betreuung der Basisstudie Schizophrenie möchte ich mich bei allen Ko-Autoren der Artikel und bei allen Studienzentren des Kompetenznetzes Schizophrenie sehr herzlich bedanken.

Und natürlich wäre diese Arbeit niemals möglich gewesen ohne die Hilfe meiner Frau Viola, die mir den nötigen Freiraum verschafft, mich immer wieder in meiner Arbeit bestärkt und in allen Entwicklungsschritten dieser Arbeit mit mir gelitten und sich mitgefreut hat. Vielen lieben Dank für diese Unterstützung!