Relapse in schizophrenia: evidence-based criteria derived by equipercentile linking and diagnostic-test-accuracy meta-analysis

Spyridon Siafis MD,[1*] Lasse Brandt MD,[2] Robert McCutcheon MD, PHD,[3,4,5,6] Stefan Gutwinski MD,[2] Johannes Schneider-Thoma MD,[1] Irene Bighelli PhD,[1] John M Kane MD,[7,8] Celso Arango MD PhD,[9,10] René S Kahn MD, PhD,[11] W Wolfgang Fleischhacker MD,[12] Patrick McGorry MD, PhD,[13,14] William T Carpenter MD,[15] Peter Falkai MD,[16] Alkomiet Hasan MD,[17] Stephen R Marder MD,[18] Nina Schooler PhD,[19] Rolf R Engel PhD,[16] William G Honer MD,[20,21] Robert W Buchanan MD,[15] Michael Davidson MD, PhD,[22,23] Mark Weiser MD, [24,25] Josef Priller MD,[1,3,26,27] John M Davis MD,[28] Oliver Howes MD, PhD, [3,4,5] Christoph U Correll MD,[7,8,29] Stefan Leucht MD[1]

[1]Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Ismaningerstr. 22, 81675, Munich, Germany.

[2]Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany.

[3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.

[4]Institute of Medical Sciences, Medical Research Council London, London, UK.

[5]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK.

[6]Department of Psychiatry, University of Oxford, Oxford, UK

[7]Department of Psychiatry, The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, USA

[8]The Donald and Barbara Zucker School of Medicine, Department of Psychiatry and Molecular Medicine, Hempstead, NY, USA.

[9]Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, School of Medicine, Universidad Complutense, Madrid, Spain.

[10]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain.

[11]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

[12]Medical University of Innsbruck, Innsbruck, Austria.

[13]Orygen, Melbourne, Australia.

[14]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia.

[15] Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, USA.

[16]Department of Psychiatry and Psychotherapy, School of Medicine, Ludwig-Maximilians-University of Munich, Munich, Germany.

[17]Department of Psychiatry, Psychotherapy and Psychosomatics, University of Augsburg, Medical Faculty, Bezirkskrankenhaus Augsburg, Augsburg, Germany.

[18]Semel Institute for Neuroscience at UCLA, VA Desert Pacific Mental Illness Research, Education and Clinical Center, Los Angeles, CA, USA.

[19]Department of Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center, Brooklyn, New York, USA.

[20]University of British Columbia, Department of Psychiatry, Faculty of Medicine, Vancouver BC, Canada

[21] BC Mental Health and Substance Use Services Research Institute, Vancouver, BC, Canada

[22]Minerva Neurosciences, Watham, MA, USA.
[23]Nicosia University School of Medicine, Nicosia, Cyprus
[24]Department of Psychiatry, Sheba Medical Center, Tel Hashomer, Israel.
[25]Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.
[26]Neuropsychiatry, Charité – Universitätsmedizin Berlin and DZNE, Berlin, Germany.
[27]University of Edinburgh and UK DRI, Edinburgh, UK.
[28]Psychiatric Institute, University of Illinois, Chicago, IL, USA.
[29]Charité Universitätsmedizin, Department of Child and Adolescent Psychiatry, Berlin, Germany.

*Corresponding author
Spyridon Siafis
Section for Evidence-Based Medicine in Psychiatry, Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Ismaningerstr. 22, 81675, Munich, Germany.
Email: spyridon.siafis@tum.de

Word count
Abstract: 251
Research in context: 431
Manuscript: 3499
References: 61
Figures: 3
Tables: 2
Supplementary files: 1

Abstract

Background

There is no consensus on the definition of relapse in schizophrenia, and scale-derived criteria with unclear clinical meaning are widely used. We therefore developed such criteria using an evidence-based, data-driven approach.

Methods

We searched the Yale University Open Data Access (YODA) database for randomized-controlled trials (RCTs) in stable patients with schizophrenia or schizoaffective disorder and obtained individual-participant-data of the Positive and Negative Syndrome Scale (PANSS), Clinical Global Impression Severity (CGI-S) and functioning scales. Change scores were linked using equipercentile linking, and PANSS-derived criteria of relapse were evaluated with diagnostic-test-accuracy meta-analysis against a clinically-relevant worsening in CGI-S ($\geq$1-point increase and score of $\geq$4) and rehospitalization.

Findings

Based on data from seven RCTs (n=2352 participants), an increase of $\geq$12 points in PANSS total had good overall test performance to identify clinically-relevant worsening in CGI-S (sensitivity=82.1%[77.0%, 86.4%], specificity=86.9%[82.9%, 90.3%]), good sensitivity but lower specify to identify rehospitalization (sensitivity=81.3%[73.6%, 87.4%], specificity: 69.3%[60.5%, 77.1%]), and corresponded to a clinically-important decline in functioning. Criteria requiring *either* an increase in PANSS total *or* in positive/disorganization symptom items of the Remission in Schizophrenia Working Group criteria had an even better performance. In contrast, the interpretation of percentage changes varied importantly across different baseline scores.

Interpretation

An increase of *either* $\geq$12-points in the PANSS total score, *or* a worsening of specific positive/disorganization symptom items could be a reasonable data-driven definition of relapse in schizophrenia, potentially linking symptoms used to define remission and relapse.

Percentage changes should not be used to define relapse because their interpretation depends on baseline scores.

Funding

German Research Foundation (grant number: 428509362)

Keywords

relapse; exacerbation; maintenance treatment; trial methodology; schizophrenia; antipsychotic; relapse prevention

Research in context

Evidence before this study

Relapse is a common occurrence in the course of schizophrenia and can be associated with negative functional outcomes and a poorer subsequent treatment response. Therefore, the prevention of relapse is a critical treatment goal, and a clear definition of relapse is paramount. However, there is currently no consensus on the definition of relapse in schizophrenia, and scale-derived criteria with ambiguous clinical relevance are widely utilized in both clinical practice and research. On April 3, 2023, we conducted a search on PubMed to identify relevant studies using the keywords "Recurrence[MeSH Terms]" AND "schizophrenia[MeSH Terms]" AND ("criteria" OR "definition*" OR "guideline*" OR "consensus") without any restrictions. Our search yielded 282 records, from which we identified attempts to operationalize the definition of relapse in schizophrenia based on expert consensus, qualitative reviews, and earlier investigations from the 1980-90s with small sample sizes. Additionally, systematic reviews, such as Moncrieff et al. (2020) in Schizophrenia Research, have highlighted the multitude of ways in which relapse can be defined in clinical trials, often with unclear clinical relevance. Consequently, the current literature highlights a pressing need for an evidence-based and data-driven definition of relapse in schizophrenia.

Added value of this study

To address this critical gap, we conducted a pioneering evaluation of the performance and clinical relevance of relapse criteria in schizophrenia, utilizing the Positive and Negative Syndrome Scale (PANSS). Through the development of operationalized criteria based on PANSS thresholds, we successfully found optimal thresholds that exhibited a remarkable good sensitivity and specificity in identifying clinically-important deterioration in the clinical global impression. Additionally, these thresholds demonstrated a commendable level of sensitivity, albeit with a slightly lower specificity, in identifying cases necessitating rehospitalization. Moreover, our findings revealed a significant challenge in interpreting

percentage changes in PANSS total scores among stable patients with schizophrenia, as the interpretation varied across different baseline scores, thus rendering it problematic in defining relapse. In contrast, absolute changes emerged as a more reliable approach for defining relapse.

Implications of the available evidence

Our evidence-based and data-driven PANSS-derived relapse criteria offer an operationalized and standardized approach to defining relapse in schizophrenia. The implementation of these criteria holds important implications for clinical research, particularly in the design of studies focused on investigating interventions for relapse prevention and examining the impact of relapse on the course of illness and long-term outcomes. Moreover, these criteria serve as a valuable tool for interpreting research findings and effectively translating them into clinical practice. The establishment of such standardized and clinically-relevant relapse criteria holds immense potential for enhancing patient care and improving overall outcomes in the management of schizophrenia.

Introduction

Relapse is common in the course of schizophrenia and may be associated with worse functional outcomes and poorer subsequent treatment response.[1,2] Therefore, relapse prevention is a major treatment goal,[3] and the definition of relapse is of core importance. However, in contrast with the consistently well-cited remission criteria of the Remission in Schizophrenia Working Group (RSWG),[4] there is a plethora of different definitions of relapse, ranging from increased service use (e.g., rehospitalization), need for treatment adjustment (e.g., dose increase), increase in psychotic symptoms (e.g., absolute and/or relative increase in total or subscale scores, increase in specific symptoms, clinical judgment), decrease in functioning, violent behavior, suicidality, and very often combinations of various criteria.[5-8]

With the exception of studies of administrative records that use hospitalization alone as a definition of relapse, changes in the severity of symptoms beyond a specified threshold are the core domains of the currently used relapse criteria, assessed with rating instruments such as the 30-item Positive and Negative Syndrome Scale (PANSS).[9] However, thresholds defining relapse vary substantially between studies (i.e., $\geq$10-points or $\geq$20-25% increase in PANSS total scores, or a worsening of selected items addressing positive symptoms), and have uncertain clinical relevance.[5-8] Data-driven approaches to defining relapse criteria in schizophrenia have not been reported.

We attempted to fill this gap using three analytic approaches: a) equipercentile linking across multiple assessment instruments,[10] b) single-group meta-analysis and c) diagnostic-test-accuracy (DTA) meta-analysis,[11] all applied to the individual-participant-data (IPD) of large scale, randomized-controlled antipsychotic drug trials (RCTs) in stable patients with schizophrenia. We used these methods to examine the clinical relevance and performance PANSS-derived criteria, and to contribute to development of evidence-based relapse criteria for future trials and studies of the implications of relapse for course of illness and long-term outcome.

Methods

We used the Yale Open Data Access (YODA) database, which enables the access and analysis of IPD from schizophrenia trials sponsored by Johnson & Johnson.[12] The meta-analysis was part of a broader project[13] and reported according to the PRISMA statement (eAppendix-1).[14] Additional methods are presented in eAppendix-2.

Dataset

Study selection

We screened the YODA database for RCTs comparing any antipsychotic medication with another and/or placebo in adult symptomatically stable (study-defined) patients with schizophrenia or schizoaffective disorder. We excluded trials in acutely-ill patients, children/adolescents, or targeting another diagnosis, e.g., insomnia.[15] All participants of the included studies were eligible, but we excluded data from a center with data integrity problems in one study.[16]

Outcome measures

We analyzed data from the following scales: PANSS (total, positive symptom original subscale[9] and Marder factor,[17] specific item scores), Clinical Global Impression change (CGI-C) and severity (CGI-S),[18] two social functioning scales, i.e., Personal and Social Performance (PSP)[19] and Social and Occupation Functioning Assessment Scale (SOFAS),[20] and psychiatric rehospitalization (study-defined).

IPD dataset

A dataset was constructed, the integrity of which was evaluated by comparing aggregated data with those presented in the primary publications.

Risk of bias

The risk of study bias was evaluated using RoB-2 [21] and QUADAS-2.[22]

Data analysis

Equipercentile linking

We investigated the relationship of change in PANSS total (absolute and percentage change) and PANSS positive symptom subscales with change in CGI-S, CGI-C and PSP/SOFAS by conducting Spearman's correlations (ρ) at last observation and equipercentile linking functions using observed cases from all available timepoints.

Equipercentile linking is a technique that identifies scores from different scales with the same percentile rank.[10,23] To aid comparisons, PSP/SOFAS scores were minus transformed so that a higher score corresponded to a poorer clinical outcome, consistent with PANSS and CGI. We explored potential subgroup differences by calculating equipercentile linking separately for each study, different timepoints, treatment groups, remission status at baseline[4] and baseline severity as measured with PANSS total.

Multiple anchors were employed: 1) worsening of $\geq$1 point in CGI-Severity as a clinically-important deterioration (=relapse),[24] 2) decline of $\geq$10 points in the PSP and SOFAS, given that these scales are originally subdivided by intervals of 10 points,[25] 3) CGI-Change of $\geq$6 ("much worse"),[24] yet important problems occurred with this approach (see eAppendix-2, results below).

Single-group random effects meta-analysis

In contrast to the above-reported rating scales, rehospitalization is an objective outcome that is frequently used as a proxy of relapse in both randomized and observational studies.[26] As this binary ('yes'/'no') outcome cannot be addressed by equipercentile linking, we conducted single-group random-effects meta-analysis to identify the mean change from baseline to the last observation on the above-mentioned scales.

IPD Diagnostic-test-accuracy meta-analysis

Our DTA meta-analysis examined the performance of operationalized PANSS-derived criteria of relapse as *index tests* to predict CGI-Severity worsening ≥1 and rehospitalization as *reference standards*. However, CGI-Severity increase ≥1 point had to be coupled with a CGI-Severity score ≥4 (moderate) to avoid increases to 2 (questionable illness) or 3 (mild),

because patients in these scenarios would still meet remission criteria and could not be considered to have relapsed.[27]

Index-tests

The index tests can be categorized into: a) PANSS total score-derived *worsening of overall symptoms* (primary index-test), b) PANSS positive subscale score-derived *worsening of positive symptoms*, c) worsening of *specific sets of PANSS positive/disorganization items* and d) multifactorial criteria requiring *either* a) *or* c). The specific PANSS item sets in b) were derived from the RSWG remission criteria[4] in an attempt to form a continuum from remission to relapse criteria. The item set from the so-called (modified) Csernansky criteria was also examined due to its frequent use in previous studies.[28,29] We evaluated various thresholds for these items in terms of severity scores (e.g., ≥4-"moderate", ≥5-"marked") and the numbers of items that worsened (e.g., ≥1 or ≥2 items). Table-1 provides an outline of the approach.

DTA meta-analysis

We calculated 2x2 contingency tables for each study and performed a two-stage random-effects DTA meta-analysis of multiple thresholds[30,31] or a bivariate DTA meta-analysis using generalized linear mixed-effects models (eApepndix-2).[32,33] We calculated sensitivity/specificity and their 95% confidence intervals (95%CI) at each threshold, presented also in a summary receiver operating characteristics (sROC) curve. We estimated optimal thresholds with the maximization of Youden's J index.[34] In order to test the robustness of the results, we conducted sensitivity analyses using alternative meta-analytic models, excluding studies without a stabilization phase or from mainland China.[35] We also explored potential subgroup differences based on the remission status at baseline [4]. The confidence of evidence was assessed with the GRADE approach.[36]

Effect-sizes for relapse

The definition of relapse may have an impact on the observed differences between antipsychotics and placebo. Therefore, we conducted random-effects meta-analysis using

odds ratios (OR),[37,38] which were also converted also absolute risks.[39] Heterogeneity was quantified using the I^2-statistic.

Data analysis was conducted in R,[40] using diagmeta,[31] equate,[41] meta,[42] and Metatron.[32]

Role of the funding source

The funders of the study had no role in the study design, data analysis, data interpretation, writing of the manuscript, or decision to publish the study findings.

Results

Description of the studies and participants

After screening 61 records, we included seven RCTs (n=2352 participants)(eAppendix-3),[16,28,29,43-46] which were conducted in multiple centers and countries (mainly Europe and the USA), except for one in mainland China.[44] Of these, five were double-blind placebo-controlled trials, investigating oral or long-acting injectable (LAI) paliperidone,[28,29,44-46] while one was open trial, comparing risperidone LAI, quetiapine and aripiprazole,[16] and another crossover trial comparing paliperidone LAI deltoid-gluteal injections,[43] with a duration ranging from 7 months (in the crossover trial) [43] to 27 months (in the open trial).[16]

The patients included in the trials were symptomatically stable and had a diagnosis of schizophrenia or schizoaffective disorder (57% male, mean age 39.5 years).[16,29] A stabilization phase preceded randomization, except for the open[16] and crossover trial.[43] Patients were required to have a PANSS total score below a defined threshold (e.g., $\leq$70-75) at baseline, except for the open trial which included patients with higher baseline scores (mean 73.0, maximum 149).[16] At baseline, the patients were on average mildly-ill (CGI-S: mean=3.1, SD=1.0; PANSS total: mean=59.2, SD=17.5), had manifest or mild difficulties in functioning (PSP: mean=71.2, SD=10.0) and 65% met criteria for symptomatic remission according to the RSWG criteria (eAppendix-4).[4] The open trial was the only providing data on CGI-Change and SOFAS,[16] while the crossover trial did not report functioning scales and rehospitalizations.[43]

The trials had some concerns regarding biases due to potential deviations from the intended interventions and outcome measurement, except for the crossover trial, which had a high risk of bias due to missing outcome data (eAppendix-5.1).[43] Additionally, there were some concerns about DTA across all trials (eAppendix-5.2)

Clinical meaning of symptom worsening

Spearman's correlations

The change scores of overall and positive symptoms in PANSS demonstrated a medium-to-large correlation with change in CGI-Severity, CGI-C, PSP, and SOFAS, as indicated by Spearman's |ρ| ranging from 0.58 to 0.78 ($p<0.001$, eAppendix-6.1).

Equipercentile linking

An increase of 1-point in CGI-S, indicating a clinically-important deterioration of global severity of illness, corresponded to an approximately 11-point worsening in PANSS total, 5-point worsening in PANSS positive subscale (original subscale and Marder factor), 10-point decline in functioning as measured by PSP and SOFAS, and CGI-C of 5.4 (i.e., midway between "minimally" and "much worse"). A larger increase of 2-points in CGI-S corresponded to an approximately 31-point increase in PANSS total, 11-point increase in PANSS positive subscale, 21-point decline in PSP/SOFAS, and CGI-C of 6.3 (eAppendix-6.2.1, Figure-1).

The results were generally robust in subgroup analyses (eAppendix-6.2.2), with one notable exception observed in the relationship between percentage changes and CGI-S. For instance, for baseline PANSS total scores between 30-45, an increase of 1-point in CGI-S corresponded to a 200% change in PANSS total, whereas for baseline scores >70, it corresponded to only 16% (Figure-2B). However, absolute changes in PANSS total were relatively consistent, with a 9-point increase for baseline scores of 30-45 and 13-point increase for baseline scores >70 corresponding to a 1-point increase in CGI-S (Figure-2A).

Mean change scores in patients readmitted to hospital

There were 172 patients readmitted to hospital (8.2% of the participants, 95%CI [7.1%, 9.4%], I^2=78.1%). On average, their global severity of illness worsened by 1.5 points ([1.2, 1.8], I^2=49.4%) in CGI-S, with CGI-C of 5.7 95%CI[5.4, 5.9], i.e., approaching "much worse". The symptoms increased with a mean of 25.6 points ([20.6, 30.7], I^2=68.9%) in PANSS total, 9.1 points ([7.8, 10.4], I^2=37%) in positive symptom subscale, while functioning

declined by a mean of 18.4 points ([14.4, 22.3], I^2=33.8%] on PSP and 11.5 points [7.8, 15.0] on SOFAS (data on SOFAS were reported only from the open trial involving more severely-ill patients at baseline).[16]

Test accuracy of PANSS-derived criteria of relapse

The performance of PANSS-derived criteria is presented in Table-2 and eAppendix-7/8.

Criteria based on worsening of overall symptoms (primary index-test)

The optimal threshold for worsening of overall symptoms was ≥12-point increase (95%[≥10, ≥14]) in PANSS total (sensitivity: 82.1%[77.0%, 86.9%], specificity: 86.9%[82.9%, 90.3%]; J=0.69) (Figure-3A) against a clinically-important deterioration of global severity of illness in CGI-S (increase ≥1 point and a score of ≥4-"moderately-ill") as the primary reference-standard,

The results were generally robust in sensitivity analyses (eAppendix-7.3), yet the overall performance was somewhat poorer in patients non-remitted (optimal threshold: ≥10 points, J=0.61) vs. remitted at baseline (optimal threshold: ≥12 points, J=0.71) (eAppendix-7.4).

When rehospitalization was used as the secondary reference-standard, the optimal threshold was ≥15-point increase (95%[≥11.0, ≥19]) in PANSS total (sensitivity: 75.8%[67.4%, 82.9%], specificity: 75.4%[67.7%, 82.0%]; J=0.51) (Figure-3B).

The confidence of the evidence at the optimal thresholds was moderate and low for the primary and secondary reference-standards, respectively (eAppendix-7.2).

Criteria based on worsening of positive symptoms

The optimal threshold for worsening of positive symptoms in PANSS positive symptom subscale was approximately ≥5-point increase when using CGI-S as the reference standard (sensitivity=78.1% [73.5%,82.1%], specificity=86% [82.8%,88.8%]; J=0.64), and ≥6-point increase when using rehospitalization (sensitivity=74.8% [66.8%,81.6%], specificity=74.8% [68.6%,80.3%]; J=0.50). Similar results were found for the positive Marder factor (eAppendix-8.1).

Criteria based on worsening of specific symptoms

The criterion "≥1 positive/disorganization RSWG item ≥1 point worse and with a score ≥4" (moderate) had good performance when compared to the primary reference-standard of a clinically-important worsening in CGI-S, with a Youden-index of J=0.68 and sensitivity and specificity being approximately the same (sensitivity=82.2%[79.0%,84.9%], specificity=85.8%[82.2%,88.7%]). The overall test performance based on the Youden-index was similar to that based on worsening of overall or positive symptoms. There were no major differences when all RSWG remission items (not just positive/disorganization), modified Csernansky criteria items, or PANSS positive symptoms were used.

When rehospitalization was used as the reference-standard the performance based on the Youden-index was generally lower. The best definition was "≥1 positive/disorganization RSWG item ≥2 points worse and a score ≥4" (sensitivity=72.1%[58.2%,82.7%], specificity=79.3[75.0%,83.0%], J=0.51). Again, criteria based on other items yielded similar results.

Multifactorial criteria based on worsening of either overall or specific symptoms

The performance of multifactorial criteria, i.e., requiring a worsening of *either* overall *or* specific symptoms, was generally better compared to their single components, with Youden-indices usually higher than 0.7. For instance, the criteria that require a worsening of either PANSS total (i.e., ≥12-point increase) or RSWG positive/disorganization items (i.e., ≥1 item, ≥1 point worse, and with a score ≥4) had a Youden index of J=0.73 when CGI-S was used as the reference standard (sensitivity=91.9% [88.7%,94.2%], specificity=80.7% [77.2%,83.7%]).

Effect-sizes for relapse

The effect-sizes for relapse between antipsychotics and placebo are presented across the different criteria in Table-2 and eAppendix-7/8. The optimal cutoffs mentioned earlier showed clear drug-placebo differences with ORs ranging from 0.25 to 0.4, which were consistent with those of the reference standards.

Discussion

We attempted to derive PANSS-based criteria for relapse in schizophrenia using individual-participant-data from 7 RCTs and 2352 patients. Our findings indicated that a worsening of ≥12 points in PANSS total, and of ≥1 point and a score ≥4 (moderate) in ≥1 positive/disorganization item of the RSWG remission criteria, corresponded to a clinically-important deterioration of ≥1 point in CGI-Severity. Requiring either one of both components further improved the overall performance. The threshold for worsening in PANSS total (i.e., ≥12-point increase) was further supported by its correspondence to a decline in functioning by ≥10 points in PSP/SOFAS. Therefore, this level of symptom severity increase in PANSS reflects an exacerbation of illness ("relapse") associated with a minimal clinically-important deterioration of global severity and decline in functioning, which may require additional intervention and support. The RSWG-based relapse criteria performed at least as well as other previously used criteria (e.g., modified Csernansky),[28,29] and may enable researchers to establish a continuum from remission to relapse with potential implications for interpreting clinical trial results.

The above PANSS-derived criteria had also a relatively good sensitivity (80-90%) in predicting rehospitalization, meaning that they are unlikely to miss relapses requiring this level of care. However, their specificity was clearly lower (60-70%). Rehospitalization is an objective and important outcome, but in contrast to PANSS, it is not well operationalized and it can vary considerably across different healthcare systems,[47,48] which may explain the relatively lower specificity. We found that the best cutoff for predicting rehospitalization was ≥15-point increase in PANSS total, but the mean PANSS total at the time of hospitalization was 24 points. This discrepancy could be explained in part by the fact that a relatively low increase in PANSS total score could predict rehospitalization earlier than the higher scores seen at the time of hospitalization. Nevertheless, as rehospitalization indicates severity and has cost implications,[47] it remains an important secondary outcome. The same may hold true for other components of previously used relapse criteria such as suicidality and violent

behavior. These are “pragmatic” criteria which can be assumed to imply a relapse, but are relatively rare and not adequately reported to allow analysis with our approach.

Previous studies have often used *percentage* change PANSS total cutoffs to define relapse,[8] perhaps consistent with the response criteria used in acute phase trials.[49-51] However, our analysis showed that such percentage cutoffs are not suitable for defining relapse in patients who are symptomatically stable at baseline. The interpretation of percentage changes varied importantly across baseline scores (Figure-2), as percentage changes are less robust when the denominators are small (i.e., baseline score minus the minimum PANSS total score of 30), which is the case in stable patients with relatively low symptom severity. To address this issue, a straightforward approach would be to employ cutoffs of absolute changes (e.g., $\geq$12-point increase in PANSS total), which have more consistent interpretation regardless of baseline severity (Figure-2).

Finally, our study found that using cutoffs of worsening in the positive symptom subscales to define relapse did not perform better compared to cutoffs of worsening in overall symptoms. This contradicts the hypothesis that criteria more focused to positive symptoms could be more precise by excluding non-specific fluctuations of mental state or withdrawal effects.[5] On the other hand, we found that a clinically-important deterioration in CGI-Severity had comparable performance to PANSS-derived criteria in predicting rehospitalizations (Table-2). However, we did find that a clinically significant worsening in the Clinical Global Impressions-Severity (CGI-S) scale had comparable performance to PANSS-derived criteria in predicting rehospitalizations (Table-2). Therefore, in low-resource settings where PANSS may not be feasible, CGI-S could be used to define relapse. This finding is further supported by a recent cohort study that demonstrated the utility of CGI-S as a transdiagnostic predictor of psychiatric hospitalization.[52]

Limitations

This analysis had certain limitations. First, we only data from the YODA database, which is limited to RCTs sponsored by Johnson & Johnson and focused mainly on risperidone and paliperidone. Therefore, we had to downrate the confidence in the evidence (eAppendix-

7.2), despite the considerable sample size of our analysis. Replications based on data from other RCTs and real-world settings are needed, including also data from antipsychotics with different pharmacological properties beyond strong postsynaptic dopamine 2 receptor antagonism,[53] as well as data regarding psychological/psychosocial treatments.[54]

Second, we employed the Youden index as a measure of the trade-off between sensitivity and specificity to determine the optimal thresholds. However, as expected, more stringent cutoffs were associated with higher specificity and lower sensitivity, and vice versa. Nevertheless, the relative effect-sizes were consistent across different scale-derived definitions of relapse, as previously demonstrated in a similar analysis of response definitions.[55] Thus, if in a given situation more sensitivity or specificity is needed, other cutoffs presented in eAppendix-7/8 may be considered.

Third, we should note that in our analysis, relapse criteria were considered fulfilled if they occurred at least once during the trial, and no time criterion was used. Nevertheless, relapses in schizophrenia may have an abrupt onset and rapid progression,[1,56] and therefore, it is important to grasp such eventual relapses. Indeed, PANSS item scores seem to increase abruptly only about one week before a relapse,[56] further highlighting the importance of identifying early warning signs that can predict timely relapses.[57-59] The development of such sensitive and timely precursors of relapse could also be useful for placebo-controlled relapse prevention trials, given the ethical and safety concerns surrounding these trials.[60] Yet, further investigation is necessary to better understand the persistence and temporal changes in symptom worsening in order to exclude potential clinically non-important short-term fluctuations of the mental state.[5]

Fourth, our study had a retrospective design and none of the trials aimed originally to investigate the performance of the relapse criteria. Therefore, different relapse criteria should also be compared prospectively in terms of their respective impact on drug-placebo difference, sensitivity/specificity against external gold standards of clinically worsening, and with regards to their effect on functioning and speed of restabilization in order to determine the optimal level of “impending relapse” that can reliably identify a clinically-relevant

worsening/relapse without exposing patients to undue short- and long-term biopsychosocial consequences of more severe relapse.

Finally, schizophrenia is a complex condition, and the performance of relapse criteria may differ across patient subgroups beyond the chronically-ill patients in the included studies. For instance, patients with a good treatment response to a first episode of schizophrenia may be truly symptom-free, and in such cases, the re-appearance of a delusion or hallucination that falls short of the level of severity defined here may still indicate a relapse.[61] At the other extreme, patients with treatment-resistant illness may live in the community with persistent disabling delusions, hallucinations and/or thought disorder (clearly not in any state of remission), yet still have periodic exacerbations of illness requiring additional interventions and a higher level of care. Our study found that the performance of relapse criteria was lower in patients who were partially remitted at baseline (eAppendix-7.4), and none of the included studies focused on treatment-resistant patients. Moreover, it is important to investigate the potential role of other illness domains apart from psychotic symptoms, such as cognitive impairment, negative symptoms and psychosocial functioning. Nonetheless, any definition of relapse must include criteria for change in symptom severity, and the definitions proposed provided a data-driven contribution to developing a comprehensive set of relapse criteria.

Conclusions

We found that *either* a $\geq$12-point increase in PANSS total score *or* worsening of $\geq$1 point and a score $\geq$4 (moderate) in $\geq$1 positive/disorganization RSWG item score can serve as an evidence-based definition of relapse with good test performance in clinically stable patients with schizophrenia. In addition, these findings may guide the development and selection of relapse criteria across a range of clinical decision points. Finally, the proposed evidence-based relapse criteria may help differentiate maintenance treatment effects of different interventions in future active- and placebo-controlled RCTs more precisely.

References

1. Emsley R, Chiliza B, Asmal L, Harvey BH. The nature of relapse in schizophrenia. *BMC Psychiatry* 2013; **13**: 50-.
2. Taipale H, Tanskanen A, Correll CU, Tiihonen J. Real-world effectiveness of antipsychotic doses for relapse prevention in patients with first-episode schizophrenia in Finland: a nationwide, register-based cohort study. *The Lancet Psychiatry* 2022.
3. Keepers GA, Fochtmann LJ, Anzia JM, et al. The American Psychiatric Association Practice Guideline for the Treatment of Patients With Schizophrenia. *American Journal of Psychiatry* 2020; **177**(9): 868-72.
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**(3): 441-9.
5. Moncrieff J, Crellin NE, Long MA, Cooper RE, Stockmann T. Definitions of relapse in trials comparing antipsychotic maintenance with discontinuation or reduction for schizophrenia spectrum disorders: A systematic review. *Schizophr Res* 2020; **225**: 47-54.
6. Olivares JM, Sermon J, Hemels M, Schreiner A. Definitions and drivers of relapse in patients with schizophrenia: a systematic literature review. *Ann Gen Psychiatry* 2013; **12**(1): 32-.
7. San L, Serrano M, Cañas F, Romero SL, Sánchez-Cabezudo Á, Villar M. Towards a pragmatic and operational definition of relapse in schizophrenia: A Delphi consensus approach. *Int J Psychiatry Clin Pract* 2015; **19**(2): 90-8.
8. Csernansky JG, Mahmoud R, Brenner R. A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *N Engl J Med* 2002; **346**(1): 16-22.
9. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin* 1987; **13**(2): 261-76.
10. Kolen MJ, Brennan RL. Test equating: Methods and practices: Springer Science & Business Media; 2013.
11. Deeks J, Bossuyt P, Leeflang M, Takwoingi Y, Flemyng E, Mellor L. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (version 2). *(under development)* 2021.
12. Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) Project--A Mechanism for Data Sharing. *N Engl J Med* 2016; **375**(5): 403-5.
13. Brandt L, Schneider-Thoma J, Siafis S, et al. Adverse events after antipsychotic discontinuation: an individual participant data meta-analysis. *Lancet Psychiatry* 2022; **9**(3): 232-42.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj* 2021; **372**: n71.
15. Luthringer R, Staner L, Noel N, et al. A double-blind, placebo-controlled, randomized study evaluating the effect of paliperidone extended-release tablets on sleep architecture in patients with schizophrenia. *Int Clin Psychopharmacol* 2007; **22**(5): 299-308.
16. Gaebel W, Schreiner A, Bergmans P, et al. Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable vs quetiapine: results of a long-term, open-label, randomized clinical trial. *Neuropsychopharmacology* 2010; **35**(12): 2367-77.
17. 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**(12): 538-46.
18. Guy W. ECDEU assessment manual for psychopharmacology: US Department of Health, Education, and Welfare, Public Health Service …; 1976.
19. Morosini PL, Magliano L, Brambilla La, Ugolini S, Pioli R. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning

Assessment Scale (SOFAS) to assess routine social funtioning. *Acta Psychiatrica Scandinavica* 2000; **101**(4): 323-9.
20. APA. Diagnostic and statistical manual of mental disorders, 4th ed. Arlington, VA, US: American Psychiatric Publishing, Inc.; 1994.
21. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* 2019; **366**: l4898.
22. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**(8): 529-36.
23. Lim RL. Linking Results of Distinct Assessments. *Applied Measurement in Education* 1993; **6**(1): 83-102.
24. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)* 2007; **4**(7): 28-37.
25. Nicholl D, Nasrallah H, Nuamah I, Akhras K, Gagnon DD, Gopal S. Personal and social functioning in schizophrenia: defining a clinically meaningful measure of maintenance in relapse prevention. *Curr Med Res Opin* 2010; **26**(6): 1471-84.
26. Taipale H, Schneider-Thoma J, Pinzón-Espinosa J, et al. Representation and Outcomes of Individuals With Schizophrenia Seen in Everyday Practice Who Are Ineligible for Randomized Clinical Trials. *JAMA Psychiatry* 2022; **79**(3): 210-8.
27. Pinna F, Deriu L, Diana E, et al. Clinical Global Impression-severity score as a reliable measure for routine evaluation of remission in schizophrenia and schizoaffective disorders. *Ann Gen Psychiatry* 2015; **14**: 6.
28. Berwaerts J, Liu Y, Gopal S, et al. Efficacy and Safety of the 3-Month Formulation of Paliperidone Palmitate vs Placebo for Relapse Prevention of Schizophrenia: A Randomized Clinical Trial. *JAMA Psychiatry* 2015; **72**(8): 830-9.
29. Fu DJ, Turkoz I, Simonson RB, et al. Paliperidone palmitate once-monthly reduces risk of relapse of psychotic, depressive, and manic symptoms and maintains functioning in a double-blind, randomized study of schizoaffective disorder. *J Clin Psychiatry* 2015; **76**(3): 253-62.
30. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology* 2016; **16**(1): 97.
31. Rücker G, Steinhauser S, Kolampally S, Schwarzer G. diagmeta: Meta-Analysis of Diagnostic Accuracy Studies with Several Cutpoints. *R package version 04-1* 2021.
32. Huang H, Huang MH. Package 'Metatron'. 2013.
33. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58**(10): 982-90.
34. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**(1): 32-5.
35. Leucht S, Li C, Davis JM, Bighelli I, Zhu Y, Furukawa TA. About the issue of including or excluding studies from China in systematic reviews. *Schizophr Res* 2022; **240**: 162-3.
36. Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *Journal of clinical epidemiology* 2020; **122**: 142-52.
37. Doi SA, Furuya-Kanamori L, Xu C, Lin L, Chivese T, Thalib L. Controversy and Debate: Questionable utility of the relative risk in clinical research: Paper 1: A call for change to practice. *Journal of Clinical Epidemiology* 2022; **142**: 271-9.
38. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Pitfalls of using the risk ratio in meta-analysis. *Res Synth Methods* 2019; **10**(3): 398-419.
39. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions: Version 6.0. Cochrane, 2019. 2019.
40. R Core Team. R: A language and environment for statistical computing. 2013.
41. Albano AD. equate: An R package for observed-score linking and equating. *Journal of Statistical Software* 2016; **74**: 1-36.

42. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evidence-based mental health* 2019; **22**(4): 153-60.
43. Hough D, Lindenmayer J-P, Gopal S, et al. Safety and tolerability of deltoid and gluteal injections of paliperidone palmitate in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 2009; **33**(6): 1022-31.
44. Rui Q, Wang Y, Liang S, et al. Relapse prevention study of paliperidone extended-release tablets in Chinese patients with schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 2014; **53**: 45-53.
45. Kramer M, Simpson G, Maciulis V, et al. Paliperidone extended-release tablets for prevention of symptom recurrence in patients with schizophrenia: a randomized, double-blind, placebo-controlled study. *J Clin Psychopharmacol* 2007; **27**(1): 6-14.
46. Hough D, Gopal S, Vijapurkar U, Lim P, Morozova M, Eerdekens M. Paliperidone palmitate maintenance treatment in delaying the time-to-relapse in patients with schizophrenia: a randomized, double-blind, placebo-controlled study. *Schizophr Res* 2010; **116**(2-3): 107-17.
47. Burns T. Hospitalisation as an outcome measure in schizophrenia. *Br J Psychiatry Suppl* 2007; **50**: s37-41.
48. Katschnig H, Straßmayr C, Endel F, et al. Using national electronic health care registries for comparing the risk of psychiatric re-hospitalisation in six European countries: Opportunities and limitations. *Health Policy* 2019; **123**(11): 1028-35.
49. Marder SR, Meibach RC. Risperidone in the treatment of schizophrenia. *Am J Psychiatry* 1994; **151**(6): 825-35.
50. Cramer J, Rosenheck R, Xu W, Henderson W, Thomas J, Charney D. Detecting improvement in quality of life and symptomatology in schizophrenia. *Schizophr Bull* 2001; **27**(2): 227-34.
51. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology* 2006; **31**(10): 2318-25.
52. Taquet M, Griffiths K, Palmer EOC, et al. Early trajectory of clinical global impression as a transdiagnostic predictor of psychiatric hospitalisation: a retrospective cohort study. *Lancet Psychiatry* 2023; **10**(5): 334-41.
53. Schneider-Thoma J, Chalkou K, Dörries C, et al. Comparative efficacy and tolerability of 31 oral and long-acting injectable antipsychotics for maintenance treatment of adults with schizophrenia: a systematic review and network meta-analysis. *Lancet* 2021; **in press**.
54. Bighelli I, Rodolico A, García-Mieres H, et al. Psychosocial and psychological interventions for relapse prevention in schizophrenia: a systematic review and network meta-analysis. *The Lancet Psychiatry* 2021; **8**(11): 969-80.
55. Furukawa TA, Akechi T, Wagenpfeil S, Leucht S. Relative indices of treatment effect may be constant across different definitions of response in schizophrenia trials. *Schizophr Res* 2011; **126**(1-3): 212-9.
56. Wang D, Gopal S, Baker S, Narayan VA. Trajectories and changes in individual items of positive and negative syndrome scale among schizophrenia patients prior to impending relapse. *NPJ Schizophr* 2018; **4**(1): 10.
57. Buck B, Hallgren KA, Campbell AT, Choudhury T, Kane JM, Ben-Zeev D. mHealth-Assisted Detection of Precursors to Relapse in Schizophrenia. *Frontiers in Psychiatry* 2021; **12**: 642200.
58. Spaniel F, Bakstein E, Anyz J, et al. Relapse in schizophrenia: definitively not a bolt from the blue. *Neuroscience letters* 2018; **669**: 68-74.
59. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J-P. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* 2018; **43**(8): 1660-6.
60. Lawrence RE, Appelbaum PS, Lieberman JA. A historical review of placebo-controlled, relapse prevention trials in schizophrenia: The loss of clinical equipoise. *Schizophr Res* 2021; **229**: 122-31.

61. Chen EY, Hui CL, Lam MM, et al. Maintenance treatment with quetiapine versus discontinuation after one year of treatment in patients with remitted first episode psychosis: randomised controlled trial. *Bmj* 2010; **341**: c4024.

Article information

Abbreviations

95%CI: 95% confidence intervals; CGI-C: Clinical Global Impression Change; CGI-S: Clinical Global Impression Severity; DTA: Diagnostic test accuracy; G1: General psychopathology item of PANSS Somatic concern; G12: General psychopathology item of PANSS Lack of judgment and insight; G14: General psychopathology item of PANSS Poor impulse control; G5: General psychopathology item of PANSS Mannerisms and posturing; G8: General psychopathology item of PANSS Uncooperativeness; G9: General psychopathology item of PANSS Unusual thought content; GRADE: Grading of Recommendations, Assessment, Development and Evaluations; I2: I-squared statistic; IPD: Individual-participant data; J: Youden's J index; LAI: Long-acting injectable; N1: Negative symptom item of PANSS Blunted affect; N4: Negative symptom item of PANSS Passive/apathetic social withdrawal; N6: Negative symptom item of PANSS Lack of spontaneity and flow of conversation; N7: Negative symptom item of PANSS Stereotyped thinking; OR: Odds ratio; P1: Positive symptom item of PANSS Delusions; P2: Positive symptom item of PANSS Conceptual disorganization; P3: Positive symptom item of PANSS Hallucinations; P4: Positive symptom item of PANSS Excitement; P5: Positive symptom item of PANSS Grandiosity; P6: Positive symptom item of PANSS Suspiciousness/persecution; P7: Positive symptom item of PANSS Hostility; PANSS: Positive and Negative Syndrome Scale; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PSP: Personal and Social Performance; QUDAS-2: Quality Assessment of Diagnostic Accuracy Studies tool version 2; RCT: Randomized-controlled trial; RoB-2Risk of Bias tool version 2; RSWG: Remission in Schizophrenia Working Group; SOFAS: Social and Occupation Functioning Assessment Scale; sROC: Summary receiver operating characteristics; USA: United States of America; YODA: Yale University Open Data Access; ρ: Spearman's correlations

Contributors

Spyridon Siafis: study concept, study design, screening of eligible studies, risk of bias evaluation, data compilation, data analysis, interpretation of the findings, writing of the first draft. Lasse Brandt: study design, data acquisition, screening for eligible studies, interpretation of the findings, revising the draft and commenting for important intellectual content. Robert McCutcheon: study design, interpretation of the findings, revising the draft and commenting for important intellectual content. Stefan Gutwinski: study design, data acquisition, interpretation of the findings, revising the draft and commenting for important intellectual content. Johannes Schneider-Thoma: risk of bias evaluation, interpretation of the findings, revising the draft and commenting for important intellectual content. Irene Bighelli: interpretation of the findings, revising the draft and commenting for important intellectual content. John M Kane: interpretation of the findings, revising the draft and commenting for important intellectual content. Celso Arango: interpretation of the findings, revising the draft and commenting for important intellectual content. René S Kahn: interpretation of the findings, revising the draft and commenting for important intellectual content. W Wolfgang Fleischhacker: interpretation of the findings, revising the draft and commenting for important intellectual content. Patrick McGorry: interpretation of the findings, revising the draft and commenting for important intellectual content. William T Carpenter: interpretation of the findings, revising the draft and commenting for important intellectual content. Peter Falkai: interpretation of the findings, revising the draft and commenting for important intellectual content. Alkomiet Hasan: interpretation of the findings, revising the draft and commenting for important intellectual content. Stephen R Marder: interpretation of the findings, revising the draft and commenting for important intellectual content. Nina Schooler: interpretation of the findings, revising the draft and commenting for important intellectual content. Rolf R Engel: interpretation of the findings, revising the draft and commenting for important intellectual content. William G Honer: interpretation of the findings, revising the draft and commenting for important intellectual content. Robert W Buchanan: interpretation of the findings, revising the draft and commenting for important intellectual content. Michael Davidson: interpretation

of the findings, revising the draft and commenting for important intellectual content. Mark Weiser: interpretation of the findings, revising the draft and commenting for important intellectual content. Josef Priller: interpretation of the findings, revising the draft and commenting for important intellectual content. John M Davis: interpretation of the findings, revising the draft and commenting for important intellectual content. Oliver Howes: study design, interpretation of the findings, revising the draft and commenting for important intellectual content. Christoph U Correll: study design, interpretation of the findings, revising the draft and commenting for important intellectual content. Stefan Leucht: project management, study concept, study design, data acquisition, risk of bias evaluation, interpretation of the findings, writing of the first draft. All authors critically reviewed the report for important intellectual content and approved the final submitted version.

Declaration of interests

Spyridon Siafis: No conflicts of interest to disclose. Lasse Brandt: No conflicts of interest to disclose. Robert A McCutcheon: He has received honoraria for educational talks from Otsuka and Janssen. Stefan Gutwinski: No conflicts of interest to disclose. Johannes Schneider-Thoma: No conflicts of interest to disclose. Irene Bighelli: No conflicts of interest to disclose. John M Kane: He has been a consultant or advisor for or has received honoraria from Alkermes, Allergan, LB Pharmaceuticals, H Lundbeck, Intracellular Therapies, Janssen Pharmaceuticals, Johnson and Johnson, Merck, Minerva, Neurocrine, Newron, Otsuka, Pierre Fabre, Reviva, Roche, Sumitomo Dainippon, Sunovion, Takeda, Teva, and UpToDate, and is a shareholder in LB Pharmaceuticals and Vanguard Research Group. Celso Arango: He has been a consultant to or has received honoraria or grants from Acadia, Abbot, Ambrosetti, AMGEN, AstraZeneca, Bristol-Myers Squibb, Dainippon Sumitomo Pharma, Forum, Gedeon Richter, Janssen Cilag, Lundbeck, Merck, Otsuka, Pfizer, Roche, Servier, Shire, Schering Plough, Sunovion and Takeda. René S Kahn: No conflicts of interest to disclose. W Wolfgang Fleischhacker: He has received consulting fees from Boehringer Ingelheim, Angelini, Richter, and Recordati, and grant support from Lundbeck

and Otsuka. Patrick McGorry: No conflicts of interest to disclose. William T Carpenter: No conflicts of interest to disclose. Peter Falkai: He received paid speakership from Otsuka, Recordati, Richter, Rovi, and Janssen and is a member of advisory boards of Richter, Rovi, and Janssen. Alkomiet Hasan: He has received paid speakerships from AbbVie, Janssen, Otsuka, Rercordati and Lundbeck. He was member of Rovi, Recordati, Otsuka, Lundbeck, and Janssen advisory boards. Stephen R Marder: He has received consulting fees from Boehringer-Ingelheim, H. Lundbeck A/S, Otsuka, Takeda, Teva, Roche, Genentech, Targacept, Forum, Abbvie, Allergan, Neurocrine. He has received research support from Boehringer-Ingelheim, Takeda, Neurocrine. Nina Schooler: He has been a consultant and/or advisor to or has received honoraria from Alkermes, Lundbeck, Otsuka, Otsuka Canada, and Teva. Rolf R Engel: No conflicts of interest to disclose. William G Honer: He has received consulting fees or sat on paid advisory boards for In Silico Biosciences, Translational Life Sciences, Newron, Otsuka and AbbVie. Robert W Buchanan: He is a DSMB member for Merck, Newron, and Roche; on the advisory board for Acadia, Boehringer Ingelheim GBMH, Karuna, Neurocrine, and Roche; and consultant for Boehringer Ingelheim GMBH. Michael Davidson: He is employee of Minerva Neurosciences. Mark Weiser: He has received advisory board, speaker's, and/or consultant fees from and owns stock with Acadia, Dexcel, Janssen, Lundbeck, Minerva, Pfizer, Roche, and Teva. Josef Priller: No conflicts of interest to disclose. John M Davis: No conflicts of interest to disclose. Oliver Howes: He was a part-time employee of H. Lundbeck A/S and has received investigator-initiated research funding from and/or participated in advisory/speaker meetings organised by Angellini, Autifony, Biogen, Boehringer-Ingelheim, Eli Lilly, Heptares, Global Medical Education, Invicro, Jansenn, Lundbeck, Neurocrine, Otsuka, Sunovion, Rand, Recordati, Roche, and Viatris/Mylan. Neither he or his family have holdings/a financial stake in any pharmaceutical company. He has a patent for the use of dopaminergic imaging. Christoph U Correll: He has been a consultant and/or advisor to or has received honoraria from: AbbVie, Acadia, Alkermes, Allergan, Angelini, Aristo, Axsome, Cardio Diagnostics, Compass, Damitsa, Gedeon Richter, Hikma, Holmusk, IntraCellular Therapies, Janssen/J&J, Karuna, LB

Pharma, Lundbeck, MedAvante-ProPhase, MedInCell, Medscape, Merck, Mindpax, Mitsubishi Tanabe Pharma, Mylan, Neurocrine, Noven, Otsuka, Pfizer, Recordati, Relmada, Reviva, Rovi, Seqirus, Servier, SK Life Science, Sumitomo Dainippon, Sunovion, Supernus, Takeda, Teva, and Viatris. He provided expert testimony for Janssen and Otsuka. He served on a Data Safety Monitoring Board for Lundbeck, Relmada, Reviva, Rovi, and Teva. He has received grant support from Janssen and Takeda. He received royalties from UpToDate and is also a stock option holder of Cardio Diagnostics, Mindpax, LB Pharma and Quantics. Stefan Leucht: In the past 3 years, SL has received honoraria as a consultant, adviser, or lecturer from Alkermes, Angelini, Eisai, Gedeon Richter, Janssen, Lundbeck, Lundbeck Institute, Merck Sharpp and Dome, Otsuka, Recordati, Rovi, Sanofi Aventis, TEVA, Medichem, and Mitsubishi.

Data sharing

Deidentified data can be analysed through the Yale University Open.

Acknowledgments

This study was funded by the German Research Foundation (grant number 428509362).

This study, carried out under YODA Project #2017-1846, used data obtained from the Yale University Open Data Access Project, which has an agreement with JANSSEN RESEARCH & DEVELOPMENT. The interpretation and reporting of research using these data are solely the responsibility of the authors and do not necessarily represent the official views of the Yale University Open Data Access Project or JANSSEN RESEARCH & DEVELOPMENT. All individual participant data used in this study can be requested from the Yale Open Data Access (YODA) project. The data request review process is described in detail on the YODA homepage.

Figure Legends

Figure-1: Equipercentile linking

Equipercentile linking of change scores using change in CGI-S as the clinical anchor. A positive score corresponds to a poorer clinical outcome (PSP and SOFAS change scores were minus transformed). The equipercentile linking between CGI-S and CGI-C can be found in the eAppendix-6.2.2. CGI-C: Clinical Global Impression Change; CGI-S: Clinical Global Impression Severity; PANSS: Positive and Negative Syndrome Scale; PANSS PS-MF: PANSS positive Marder factor; PANSS PS-S: PANSS positive symptom subscale; PSP: Personal and Social Performance; SOFAS: Social and Occupational Functioning Assessment Scale.

Figure-2: Linking of absolute and percentage change scores in PANSS total and baseline severity

Equipercentile linking between absolute change (Figure-2A) and percentage change (Figure-2B) in PANSS total using change in CGI-S as the clinical anchor. A positive score corresponds to a poorer clinical outcome. There was some variability in the linking functions of absolute change scores across different baseline severity levels, e.g., an increase of 1-point in CGI-S corresponded to a range of increase in PANSS total from 9 (baseline PANSS total of 30-45) to 13 (baseline PANSS total >70), an increase of 2-points in CGI-S to a range from 25 to 34 and of 3-points to a range from 47 to 55. This variability was substantial and more pronounced in the linking functions of percentage change scores, e.g., an increase of 1-point in CGI-S corresponded to a range of percentage increase in PANSS total from 16% (baseline PANSS total >70) to 200% (baseline PANSS total 30-45), an increase of 2-points in CGI-S to a range from 50% to 340% and of 3-points to a range from 88% to >400%. CGI-S: Clinical Global Impression Severity, PANSS: Positive and Negative Syndrome Scale

Figure-3: sROC curve for criteria based on worsening of overall symptoms (primary test index)

Summary receiver operating curves (sROC) curves of the DTA meta-analysis of multiple thresholds of criteria based on worsening of overall symptoms as measured with change PANSS total against two reference standards (A. CGI-S criterion of relapse, and B. rehospitalization. Each point represents the sensitivity and false positive rate (1-specificity) for each of the seven studies (shown with a different key color) across different cutoffs of absolute change in PANSS total. The study of Hough et al 2009 did not provide data for rehospitalizations. The solid black line represents the sROC curve with the 95%CI for sensitivity and specificity. The cross represents the summary point given the cutoff that maximized the Youden index (about $\geq$12, 95%[$\geq$10, $\geq$14] points increase in PANSS total in A and $\geq$15 95%[$\geq$11, $\geq$19] in B) and the ellipse its confidence interval. The crossover study was excluded from the analysis using rehospitalizations, since they were not systematically assessed and no relapse criteria were used.[43] CGI-S: Clinical Global Impression Severity; DTA: Diagnostic test accuracy; PANSS: Positive and Negative Syndrome scale, sROC: summary receiver operating characteristics curve, 95%CI: 95% confidence intervals.

Tables

Table 1 Operationalized criteria of relapse used in the study

	Domain	Definition
Primary index-test	**Worsening of overall symptoms**	Increase from ≥1 to ≥30 points in PANSS total score
Secondary index-test	**Worsening of positive symptoms**	a. Increase from ≥1 to ≥15 points in PANSS positive symptom subscale score b. Increase from ≥1 to ≥15 points in PANSS positive Marder factor score
	Worsening of specific symptoms	a. Increase ≥1-2 points and a score ≥4-5 in ≥1-2 items for positive and disorganization symptoms in the RSWG criteria (P1, P2, P3, G5, G9) b. Increase of ≥1-2 points and a score ≥4-5 in ≥1-2 items for positive, negative and disorganization symptoms in the RSWG criteria (P1, P2, P3, N1, N4, N6, G5, G9) c. Increase of ≥1-2 points and a score ≥4-5 in ≥1-2 items in the modified Csernansky criteria (P1, P2, P3, P4, P6, P7, G8, G14) d. Increase of ≥1-2 points and a score ≥4-5 in ≥1-2 items for positive symptoms in the original positive symptom subscale (P1, P2, P3, P4, P5, P6, P7) e. Increase of ≥1-2 points and a score ≥4-5 in ≥1-2 items for positive symptoms in the positive Marder factor (P1, P3, P5, P6, N7, G1, G9, G12)
	Multifactorial criteria	Requiring *either* a) a worsening of overall symptoms as measured with ≥10, ≥12 or ≥15 points increase in PANSS total score *or* b) a worsening of specific symptoms (as defined above)
Primary reference-standard	**Clinically important worsening of the global severity of illness**	Increase of ≥1-point and a score ≥4 in CGI-S
Secondary reference-standard	**Rehospitalization**	Psychiatric rehospitalization

A patient was classified as relapsed when fulfilled the operationalized criteria at least once during the trial. CGI-S: Clinical Global Impression Severity scale; PANSS: Positive and Negative Syndrome Scale; PANSS items P1: Delusions; P2: Conceptual disorganization; P3: Hallucinations; P4: Excitement, P5: Grandiosity; P6: Suspiciousness/persecution; P7: Hostility; N1: Blunted affect; N4: Passive/apathetic social withdrawal; N6: Lack of spontaneity and flow of conversation, N7: Stereotyped thinking, G1: Somatic concern; G5: Mannerisms and posturing; G8: Uncooperativeness; G9: Unusual thought content; G12: Lack of judgment and insight; G14: Poor impulse control; RSWG: Remission in Schizophrenia Working Group

Table 2 Test performance of PANSS-derived criteria of relapse

Index test		Reference standard: CGI-S increase of ≥1 point and ≥4			Reference standard: rehospitalization*			Effect-sizes for relapse between antipsychotics and placebo (%, OR)
		Sensitivity [95%CI]	Specificity [95%CI]	J index	Sensitivity [95%CI]	Specificity [95%CI]	J index	
Cutoffs of increase in PANSS total	≥5-point increase in PANSS total	92.7% [88.8%,95.5%]	68.6% [60%,76.4%]	0.61	90.8% [85%,94.8%]	53% [42.3%,63.4%]	0.44	42% vs. 62%, OR=0.45 [0.36,0.56], I^2=14.3%
	≥10-point increase in PANSS total	85.8% [81%,89.7%]	82.7% [77.4%,87.1%]	0.68	84.5% [77.3%,90%]	64.8% [55.5%,73.4%]	0.49	23% vs. 50%, OR=0.3 [0.23,0.39], I^2=17.2%
	≥12-point increase in PANSS total	82.1% [77%,86.4%]	86.9% [82.9%,90.3%]	0.69	81.3% [73.6%,87.4%]	69.3% [60.5%,77.1%]	0.51	19% vs.45%, OR=0.29 [0.2,0.42], I2=51.5%
	≥15-point increase in PANSS total	75.4% [70.1%,80.2%]	91.9% [89.2%,94%]	0.67	75.8% [67.4%,82.9%]	75.4% [67.7%,82%]	0.51	15% vs. 40%, OR=0.26 [0.17,0.4], I^2=54.6%
	≥20-point increase in PANSS total	62% [56.3%,67.4%]	96.8% [95.6%,97.7%]	0.59	65% [55.6%,73.5%]	84% [78%,88.8%]	0.49	10% vs. 31%, OR=0.25 [0.18,0.34], I^2=12.6%
	≥25-point increase in PANSS total	46.9% [40.7%,53.1%]	98.9% [98.4%,99.3%]	0.46	52.7% [42.7%,62.6%]	90.4% [85.8%,93.7%]	0.43	7% vs. 22%, OR=0.26 [0.18,0.38], I^2=0%
Cutoffs of worsening in specific items	≥1-point increase and score of ≥4 in ≥1 items of positive/disorganization symptoms in RSWG (P1, P2, P3, G5, G9)	82.2% [79%,84.9%]	85.8% [82.2%,88.7%]	0.68	80.8% [65.9%,90.2%]	69.3% [65.3%,73.1%]	0.50	24% vs. 43%, OR=0.43 [0.32,0.57], I^2=34%
	≥2-point increase and score of ≥4 in ≥1 item of positive/disorganization symptoms in RSWG (P1, P2, P3, G5, G9)	70.3% [66.2%,74.1%]	93.7% [92.1%,95%]	0.64	72.1% [58.2%,82.7%]	79.3% [75%,83%]	0.51	14% vs.35%, OR=0.31 [0.24,0.41], I^2=0%
	≥1-point increase and score of ≥4 in ≥1 item of Csernansky criteria (P1, P2, P3, P4, P6, P7, G8, G14)	89% [86.4%,91.2%]	81.9% [76.9%,86.1%]	0.71	83.5% [77.2%,88.4%]	65.4% [61.3%,69.2%]	0.49	28% vs. 47%, OR=0.43 [0.32,0.58], I^2=35.5%
	≥2-point increase and score of ≥4 in ≥1 item of Csernansky criteria (P1, P2, P3, P4, P6, P7, G8, G14)	78.9% [75.5%,81.9%]	89.9% [86.3%,92.6%]	0.69	78.1% [67.7%,85.9%]	74% [69.3%,78.1%]	0.52	18% vs.40%, OR=0.32 [0.25,0.42], I^2=0%
Multifactorial criteria: cutoffs of	≥12-point increase in PANSS total *or* ≥1-point increase and score of ≥4 in ≥1 items of positive/disorganization symptoms	91.9% [88.7%,94.2%]	80.7% [77.2%,83.7%]	0.73	90.9% [75.8%,96.9%]	62.5% [56.7%,67.9%]	0.53	32% vs.53%, OR=0.41 [0.31,0.54], I^2=38.5%

increase in PANSS total or cutoffs of worsening of specific items	in RSWG (P1, P2, P3, G5, G9)							
	≥15-point increase in PANSS total *or* ≥2-point increase and score of ≥4 in ≥1 item of positive/disorganization symptoms in RSWG (P1, P2, P3, G5, G9)	84.6% [79.2%,88.7%]	90% [88.4%,91.4%]	0.75	84.4% [71.2%,92.2%]	72% [65%,78%]	0.56	20% vs.45%, OR=0.31 [0.24,0.39], I2=0%
	≥12-point increase in PANSS total *or* ≥1-point increase and score of ≥4 in ≥1 item of Csernansky criteria (P1, P2, P3, P4, P6, P7, G8, G14)	94.1% [91.3%,96%]	77.4% [73.6%,80.9%]	0.71	91.1% [78.3%,96.7%]	60% [54.7%,65.1%]	0.51	33% vs.55%, OR=0.41 [0.33,0.51], I^2=21.3%
	≥15-point increase in PANSS total *or* ≥2-point increase and score of ≥4 in ≥1 item of Csernansky criteria (P1, P2, P3, P4, P6, P7, G8, G14)	87.3% [82.9%,90.7%]	87% [84.2%,89.3%]	0.74	86.8% [73.1%,94.1%]	68.9% [62.5%,74.6%]	0.56	22% vs.48%, OR=0.31 [0.25,0.4], I^2=0%
Clinically-important worsening in CGI-Severity	≥1-point increase in CGI-Severity and a score of ≥4	n.a.			79.5%[70.7%, 86.1%]	75.8%[70.0%, 80.8%]	0.55	17% vs. 39%, OR=0.33 [0.29, 0.37], I^2=0%

*The crossover study was excluded from the analysis using rehospitalizations, since they were not systematically assessed and no relapse criteria were used [43]. The results for the rest of the cutoffs are presented in the eAppendix-7/8. The effect-sizes for rehospitalizations were 3% vs. 8%, OR=0.35 [0.27, 0.45], I^2=0%. CGI-S: Clinical Global Impression Severity scale; I2=I-squared; OR: Odds ratio; PANSS: Positive and Negative Syndrome Scale; 95%CI: 95% confidence intervals; P1: Delusions; P2: Conceptual disorganization; P3: Hallucinations; P4: Excitement, P5: Grandiosity; P6: Suspiciousness/persecution, P7: Hostility, N1: Blunted affect; N4: Passive/apathetic social withdrawal; N6: Lack of spontaneity and flow of conversation, N7: Stereotyped thinking, G1: Somatic concern; G5: Mannerisms and posturing; G8: Uncooperativeness; G9: Unusual thought content; G12: Lack of judgment and insight; G14: Poor impulse control. RSWG: Remission in Schizophrenia Working Group; n.a.: not applicable.