Both scales were administered by a cohort of remote centralized clinicians who were blinded to protocol requirements, and aggregated data from seven global clinical trials of schizophrenia in which both the PANSS and CGI-S were administered. The objectives of the present study were to:

1. Replicate and expand upon Leucht et al. using a large dataset spanning multiple clinical trials.
2. Examine the level of agreement when both measures are administered by the same centralized and calibrated clinical trial raters.

The objectives of the present study were to:

• Aggregated data from seven global clinical trials of schizophrenia in which both the PANSS and CGI-S were administered.
• Both scales were administered by a cohort of remote centralized clinicians who were blinded to protocol requirements, visit number, and prior knowledge of subject, a methodology that has been shown to help standardize assessments and eliminate scoring bias.
• The centralizers completed the PANSS and CGI-S ratings, after which they examined alignment between the scores according to the guideline provided by Leucht et al. Misaligned scores were noted but initial scores remained unchanged.

METHODS

• The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the usefulness of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

CONCLUSIONS

• The present study utilized a large dataset spanning multiple global schizophrenia studies, replicated and expanded upon the original 2005 work of Leucht and colleagues, and provided additional evidence for substantial agreement between PANSS total score and CGI-S.
• This study, however, found much higher correlations (r = .91, p < .0001), likely attributable to implementability of assessments by highly calibrated central clinicians.
• As noted in the methods, raters were blinded to the study protocol, treatment arm and visit, which appeared to reduce bias (therapeutic alliance, expectancy, etc.) that often occurs in clinical trials and can compromise ratings.
• Less bias and associated variance translates to potentially increased power to detect a therapeutic effect.
• The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the usefulness of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias, ceiling effects, and lack of data with regards to reliability and validity.

Figure 1. Relationships between PANSS and CGI-S scores

<table>
<thead>
<tr>
<th>Table 2. Correlations between PANSS total scores and CGI-S scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>Pearson Correlation Coefficient</td>
</tr>
</tbody>
</table>

# Subjects 262 473 738 334 418 422 423 3,070
# PANSS/CGI-S Assessments 1,031 2,068 3,583 1,581 1,338 1,900 975 12,476
Mean PANSS (SD) 82.2 (15.73) 81.9 (14.93) 79.9 (14.0) 80.2 (15.42) 70.1 (15.99) 88.2 (16.66) 90.8 (13.88) 81.5 (16.01)
Mean CGI-S (SD) 4.4 (0.88) 4.3 (0.83) 4.2 (0.76) 4.3 (0.87) 3.7 (0.95) 4.7 (0.91) 4.8 (0.75) 4.3 (0.85)

Misalignments were very low, and observed in only 1,159 (9%) of PANSS/CGI-S scores.

References
1. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS): a decimal Global Impression-Score (CGI) scale are frequently used as endpoints in schizophrenia clinical trials.
2. Although both measures are clinician-reported outcome measures (ClinRO), the PANSS examines the schizophrenia symptomatology and general psychopathology, whereas the CGI provides a global assessment of overall functioning.
3. The CGI, comprised of three items, measures illness severity (CGI-S) and improvement (CGI-I), and contains a single item related to drug treatment effects (i.e., efficacy index).
4. Each CGI item is designed to be rated separately, there is no composite score.
5. Clinicians are permitted access to all collateral patient information and expected to rate subjects relative to their past experience with other schizophrenia patients.
6. Many clinical trials require that the CGI be administered last to consider all collected information when making the global rating(s).

The CGI, however, has been criticized for its psychometric limitations4-7 and susceptibility to expectancy bias when administered following the PANSS3.

It also observed a correspondence between the two scores, where mildly ill (CGI-S 3) corresponded to a total PANSS score of 81.5. (SD=16) and 4.3 (SD =0.9) for the PANSS total and CGI-S scores, respectively. This corresponds to a ‘moderately’-to- ‘markedly ill’ aggregate population8.

The mean (SD) PANSS total scores at each CGI severity stage are shown in Figure 1.

The distributions of PANSS total scores at each of the seven stages all fell within range of the Leucht et al. findings: 37.2 (4.3), 47.1 (5.8), 59.9 (5.9), 76.4 (6.6), 92.9 (6.7), 109.4 (8.6), and 136.3 (12.7), respectively.

Importantly, we observed a strong and significant correlation between PANSS total and CGI-S scores, r = .91, p < .0001. The Pearson correlation coefficients for each of the studies within the cohort are shown in Table 2, and ranged from .80 to .92 (versus 0.56 to 0.73 reported by Leucht et al.3).

The mean (SD) CGI-S total scores at each of the seven stages all fell within range of the Leucht et al. findings: 37.2 (4.3), 47.1 (5.8), 59.9 (5.9), 76.4 (6.6), 92.9 (6.7), 109.4 (8.6), and 136.3 (12.7), respectively.

The distributions of PANSS total scores at each CGI severity stage are shown in Figure 1.

The present study also had limitations, particularly those inherent in aggregate data, such as selection of studies and potential bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the usefulness of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias, ceiling effects, and lack of data with regards to reliability and validity.

The present study also had limitations, particularly those inherent in aggregate data, such as selection of studies and potential bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

The centralizers completed the PANSS and CGI-S ratings, after which they examined alignment between the scores according to the guideline provided by Leucht et al. Misaligned scores were noted but initial scores remained unchanged.

Conclusions

The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the usefulness of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias, ceiling effects, and lack of data with regards to reliability and validity.

The present study also had limitations, particularly those inherent in aggregate data, such as selection of studies and potential bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the usefulness of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias, ceiling effects, and lack of data with regards to reliability and validity.

The present study also had limitations, particularly those inherent in aggregate data, such as selection of studies and potential bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

CONCLUSIONS

The present study utilized a large dataset spanning multiple global schizophrenia studies, replicated and expanded upon the original 2005 work of Leucht and colleagues, and provided additional evidence for substantial agreement between PANSS total score and CGI-S.

This study, however, found much higher correlations (r = .91, p < .0001), likely attributable to implementability of assessments by highly calibrated central clinicians.

As noted in the methods, raters were blinded to the study protocol, treatment arm and visit, which appeared to reduce bias (therapeutic alliance, expectancy, etc.) that often occurs in clinical trials and can compromise ratings.

Less bias and associated variance translates to potentially increased power to detect a therapeutic effect.

The substantial overlap and significant shared variance between the two scales (i.e., 83 percent) brings into question the additional value of CGI-S administration, as this scale has already been shown to have several limitations, including expectancy bias, ceiling effects, and lack of data with regards to reliability and validity.

The present study also had limitations, particularly those inherent in aggregate data, such as selection of studies and potential bias in the design of individual studies. It is noteworthy, however, that despite such limitations, correlations remained robust.

Considering the complementary nature of the PANSS and CGI-S and the significant correlations between these measures, researchers should carefully consider the risk-benefit of including both on the chance they do not align and/or consider implementing a blinded centralized review (over-read) of all collected data.