



Site-independent confirmation of primary site-based PANSS ratings in a schizophrenia trial

Steven D. Targum^{a,*}, Christopher Murphy^a, Alan Breier^b, Stephen K. Brannan^c

^a Signant Health, Boston, MA, USA

^b Indiana University School of Medicine, Indianapolis, IN, USA

^c Karuna Therapeutics, Boston, MA, USA

ARTICLE INFO

Keywords:

PANSS
Schizophrenia
Paired ratings
Surveillance
Functional unblinding

ABSTRACT

Blinded, site-independent (remote) ratings from audio-digital recordings of site-based Positive and Negative Syndrome Scale (PANSS) interviews were obtained in a 5-week, randomized, double-blinded study assessing the safety, tolerability, and efficacy of KarXT (a fixed combination of xanomeline and trospium chloride) in hospitalized adults with schizophrenia experiencing an acute exacerbation of psychosis (EMERGENT-1; [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03697252) identifier: NCT3697252). The blinded site-independent raters had no knowledge of site location, study visit, drug vs. placebo assignment, or any treatment emergent adverse events (TEAEs). Concordance analyses of 561 paired site-based and site-independent PANSS ratings across all visits revealed a high correlation (ICC = 0.775). Paired scoring differences were positively correlated with the PANSS total score (Spearman's rho = 0.37, $p < 0.0001$). Paired PANSS scores were available from 148 subjects at both the baseline and end of study visits (KarXT = 72, Placebo = 76). Site-based PANSS total scores (primary aim) revealed a significantly greater improvement from baseline in the KarXT group compared to the placebo group ($p < 0.0001$). The blinded site-independent PANSS total scores derived from listening to and scoring the recorded site-based PANSS interviews replicated this finding ($p < 0.001$) and yielded an overall predictive value of 85.1% for matching the site-based response/non-response outcomes. TEAEs have the potential to “unblind” site-based ratings. In this study, the site-independent raters were blinded to TEAEs, affirmed the site-based PANSS ratings, and mitigated concerns about possible functional unblinding of site-based raters. This method of blinded assessment via audio-digital recordings may have utility for other studies concerned with ratings precision and/or functional unblinding.

1. Introduction

The use of blinded, site-independent raters to review interview quality and score audio-digitally recorded site-based subject interviews (“paired” ratings) is a quality assurance (QA) surveillance strategy that has been effectively used to monitor and assess ratings reliability in clinical trials (Targum et al., 2014, 2015, 2019, 2021; Targum and Catania, 2019). Site-independent raters have access to the recorded information and digital notes obtained during the interview, but are blinded to the study site location, study visit, and drug assignment. Consequently, site-independent raters are essentially blinded to any treatment-emergent adverse events (TEAEs) that may occur during the study and their data can be used to independently assess site-based ratings without concerns about functional unblinding. We applied this blinded assessment method to compare the data obtained from paired

site-based rater interviews during a clinical trial of hospitalized adults with schizophrenia (the EMERGENT-1 trial).

The Positive and Negative Syndrome Scale (PANSS) has been the standard instrument used to assess symptomatic change in clinical studies of schizophrenia (Kay et al., 1987). Reliable administration of the PANSS requires competent, well trained and consistent interviewers, reliable informant information, and a cooperative subject from visit to visit. The use of multiple raters at multiple trial sites introduces variance that can reduce inter-rater reliability, impact data quality, and adversely affect trial outcome (Muller and Szegedi, 2002; Berendsen et al., 2020). It has been shown that the use of audio-recording of site-based clinical interviews with site-independent review can improve ratings precision (Targum et al., 2014, 2015, 2021; Targum and Catania, 2019).

We conducted a paired ratings analysis of the PANSS during a 5-week, phase 2, randomized, double-blinded trial that assessed the

* Corresponding author. 505 Tremont Street, Unit 907, Boston, MA, 02116, USA.

E-mail address: sdtargum@yahoo.com (S.D. Targum).

<https://doi.org/10.1016/j.jpsychires.2021.10.027>

Received 14 June 2021; Received in revised form 10 October 2021; Accepted 19 October 2021

Available online 21 October 2021

0022-3956/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

efficacy, safety, and tolerability of KarXT (a fixed combination of xanomeline and trospium) in hospitalized adults with schizophrenia experiencing an acute exacerbation of psychosis (EMERGENT-1; [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03697252) identifier: NCT03697252). Xanomeline tartrate, an M1/M4-preferring agonist of muscarinic acetylcholine receptors, was initially developed during the 1990s to treat Alzheimer's disease, and showed dose-dependent improvement in psychotic symptoms (Bodick et al., 1997). However, its development for this disorder was curtailed due mostly to gastrointestinal side effects such as vomiting, diarrhea, and orthostasis attributed to muscarinic receptor agonism (Shannon et al., 1994; Bodick et al., 1997; Messer, 2002). Subsequently, a small double-blind, placebo-controlled study in 20 participants with schizophrenia showed a marked improvement in psychotic symptoms following xanomeline treatment that has compelled further investigation of this compound (Shekhar et al., 2008). Trospium chloride is a peripheral anticholinergic that is marketed for the treatment of overactive bladder that does not cross the blood-brain barrier (Pak et al., 2003). The addition of trospium to xanomeline reduced the pro-cholinergic adverse effects associated with xanomeline by about 50% in a Phase 1 trial (Kavoussi et al., 2017; Brannan et al., 2018). The EMERGENT-1 study sought to ameliorate the TEAEs associated with xanomeline by blocking the activation of peripheral muscarinic receptors with trospium in the treatment of acutely psychotic subjects with schizophrenia.

The results of the EMERGENT-1 study revealed that subjects assigned to KarXT did significantly better than placebo-assigned subjects on the total PANSS scores, the positive and negative syndrome scales, and the Marder factors (Brannan et al., 2021). In the analyses reported here, we compared the total PANSS score at baseline and week 5 (or early termination visit) obtained by the site-based raters with the paired scores of the blinded, site-independent raters to affirm the primary site-based PANSS outcomes and to mitigate concerns about possible functional unblinding.

2. Methods

Data derived for this paired ratings analysis came from the EMERGENT-1 clinical trial. The primary objective of the study was to assess the efficacy of KarXT versus placebo in reducing PANSS total scores (primary aim) in adult inpatients with a diagnosis of schizophrenia. The study was sponsored by Karuna Therapeutics Inc. and was conducted between September 2018 and September 2019 at 12 study centers in the United States. The study was conducted in accordance with the Declaration of Helsinki (1964) and Good Clinical Practices as outlined by the International Conference on Harmonisation (1997). All subjects consented to study participation and to audio recording of key interviews (described below). The full results of this study have been reported previously (Brannan et al., 2021).

2.1. Study eligibility and design

Eligible subjects were men or women between the ages of 18 and 60 (inclusive) who met DSM5 criteria for schizophrenia confirmed by the Mini-International Neuropsychiatric Interview (M.I.N.I.) and presented with an acute exacerbation of psychosis or a relapse of symptoms with an onset less than 2 months before the screening visit (Sheehan et al., 1998; APA, 2013; Brannan et al., 2021). The inclusion criteria required a PANSS total score between 80 and 120 (inclusive), an item score ≥ 4 on at least 2 of 4 key positive symptom items (delusions, conceptual disorganization, hallucinatory behavior, suspiciousness), and a Clinical Global Impression of severity (CGI-S) score ≥ 4 at the screen and baseline visits (Haro et al., 2003). Additional full details of the study design and subject eligibility requirements have been described elsewhere (Brannan et al., 2021).

Safety and tolerability were assessed based on spontaneously reported adverse events and by measuring vital signs, clinical laboratory

evaluations, ECG parameters, and suicidal ideation with the Columbia Suicide Severity Rating Scale (Posner et al., 2011). Treatment-emergent adverse events (TEAEs) were rated for severity and summarized by treatment group. An adverse event was defined as any untoward medical occurrence in a clinical study subject and could therefore be any unfavorable and unintended sign, symptom, or disease temporally associated with the use of the investigational product (KarXT) or placebo whether or not it was related to the investigational product.

Enrolled subjects were randomized to either oral KarXT or placebo in a 1:1 ratio for a treatment period of 5 weeks. KarXT was titrated to a maximum dose of xanomeline 125 mg/trospium 30 mg BID. Flexible dosing was employed to maximize therapeutic benefit while avoiding intolerable adverse events. Dosing began with a lead-in dose of xanomeline 50 mg with trospium 20 mg BID for the first 2 days followed by xanomeline 100 mg/trospium 20 mg BID for the remainder of Week 1 (Days 3–7), and titrated upwards on Day 8 to xanomeline 125 mg/trospium 30 mg BID unless the subject was continuing to experience adverse events from the previous dose increase.

Signant Health was contracted to conduct a quality assurance (QA)-monitoring program of site-based PANSS ratings at all visits. The QA program assessed interview quality and scoring concordance between paired site-based and site-independent PANSS ratings to assess ratings reliability of the primary measure and identify ratings "outliers" with discordant scores that required remediation during the study. Rater remediation involved a 1:1 discussion with the identified rater about interview administration and/or scoring conventions as needed (Targum et al., 2014, 2015).

2.2. Audio-digital recording method of site-independent PANSS assessment

The site-based PANSS interviews were administered using the SCI-PANSS (Structured Clinical Interview for the assessment of Positive and Negative Syndrome Scale, Multi-Health Systems, Inc). All 27 site-based and 3 site-independent raters were physicians and participated in a comprehensive rater training and certification program that included didactic presentations, observation of expert PANSS video interviews, and demonstration of PANSS scoring competency via inter-rater reliability (IRR) assessments of the PANSS video interviews.

The site-independent PANSS ratings were derived from audio recordings of the site-based PANSS interviews, with accompanying digital notes containing corroborative informant information completed by the site-based raters. Some PANSS items require direct observation of subject behavior during the interview and cannot be adequately assessed by listening to a recorded interview even with digital notes. In this study, ratings for items N1 (blunted affect), G4 (tension), G5 (mannerisms and posturing), and G7 (motor retardation) were carried over from the site-based score. The concordance analysis algorithm for calculation of the total independent PANSS score accepted the site-based score for 4 of the 30 items.

Every PANSS interview at every study visit was audio-digitally recorded and electronically transmitted to Signant Health for QA (surveillance) review. Upon receipt and preliminary review of audio quality, a pre-determined percentage of interviews across all study visits were sent to one of the 3 blinded PANSS reviewers for independent scoring based upon the audio recording and available digital notes. The digital notes contained corroborative information as written by the site-based raters during the interviews, but no other information about other scales or possible adverse events. The surveillance program planned to review all of the PANSS interviews that were conducted at the baseline and endpoint visits. In this way, it was possible to compare the site-independent results with the primary site-based rater's results.

2.3. Statistical analyses

Data for this report used recorded site-based PANSS interviews

paired with site-independent PANSS reviewer scores where the transmission of the audio-digital recording afforded a complete PANSS interview for satisfactory site-independent review. The primary endpoint of the EMERGENT-1 trial was the least-squares mean change from baseline on the PANSS total score at week 5. The modified intent to treat (mITT) population was used for all efficacy analyses and included all subjects who were randomized, received at least one dose of study medication, had a baseline and at least one postbaseline PANSS assessment. The majority of PANSS interviews at the baseline and endpoint/early termination visits were recorded and independently scored. Therefore, the analysis in this report included nearly all of the mITT population that had available pairs of PANSS assessments at baseline and endpoint/early termination visits. No data points were imputed in these analyses. Statistical evaluation of the data included analysis of covariance (ANCOVA), paired t tests, and chi square analyses using SAS v9.4 software (SAS Institute Inc, 2013).

This comparative analysis focused on the baseline and endpoint/early termination visits. We used ANCOVA to examine the change from baseline for the total PANSS, positive and negative symptom subscales and the 4 key positive symptoms that were part of the inclusion criteria (delusions, conceptual disorganization, hallucinatory behavior, suspiciousness). We used $\geq 30\%$ PANSS total score improvement from baseline to endpoint as the criterion for treatment response and subtracted 30 points from the baseline PANSS score due to PANSS scaling from 1 to 7 for each of the 30 items as opposed to 0–6 (Leucht et al., 2007, 2010; Obermeyer et al., 2010; Correll et al., 2011).

3. Results

There were 170 enrolled subjects who received at least one dose of study medication and had at least one postbaseline PANSS assessment (modified ITT). Overall, the mean age of the mITT subjects was 42.8 ± 10.0 (SD) years, the mean BMI was 29.1 ± 5.2 , and 131 subjects (77.1%) were male. There were no meaningful demographic differences between the subjects assigned to KarXT or placebo.

Concordance analyses between the paired site-based and site-independent ratings across all visits were conducted on the audio-recorded interviews with acceptable audio transmission. There were

561 paired PANSS scores available for analysis across all visits. Within the mITT population, there were 148 paired PANSS ratings with both baseline and endpoint (or early termination) visits available for treatment outcome analysis and comparison between the site-based and site-independent ratings. The demographics of the paired ratings group did not differ from the mITT population: the mean age was 42.3 ± 10.2 years, and 112 subjects were male (75.7%). Within this group, 133 subjects completed the study, and 15 subjects withdrew before week 5 but had at least one post-baseline PANSS assessment (an early termination visit) in which PANSS assessments were obtained.

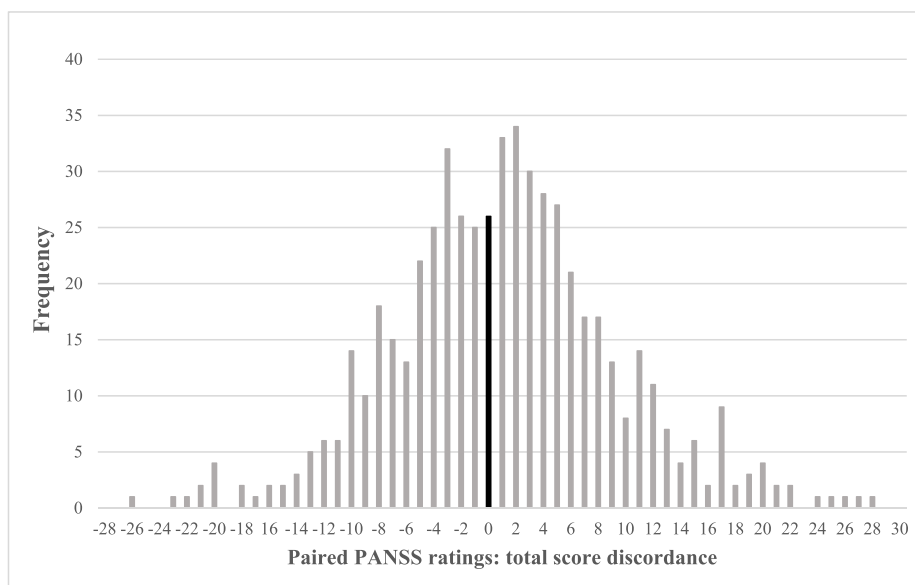
3.1. Concordance analyses between site-based and site-independent PANSS ratings

The mean site-based PANSS total score across all study visits for the 561 paired PANSS scores was 93.0 ± 12.8 (SD) and the mean paired site-independent PANSS scores was 91.9 ± 12.1 ($t = 3.15$; $df = 560$; $p < 0.002$). Fig. 1 displays the paired scoring difference distribution between the site-based and site-independent PANSS scores. The intra-class correlation (ICC) between site-based and paired independent total PANSS scores was 0.775 ($p < 0.0001$). Of the 561 paired scores, 308 (54.9%) differed by ≤ 5 points, 454 (80.9%) differed by ≤ 10 points, and 14 pairs (2.5%) had > 20 points scoring difference.

The QA program identified and remediated ratings “outliers” during the study. The higher paired scoring discordances were due either to administrative error (data entry), poor audio quality, lack of detailed digital notes, or poor interview quality affecting independent replication. Following efforts at rater remediation, 2 raters were dismissed from the study, whereas the other raters demonstrated improvement during the remainder of the study.

3.2. Effect of symptom severity on paired scoring discordance

Paired scoring differences were positively correlated with the total PANSS score (Spearman’s $\rho = 0.37$, $p < 0.0001$). There was a bidirectional pattern of paired scoring differences that was related to the magnitude of symptom severity. Total site-based PANSS scores ≥ 100 generated statistically significant positive mean discordance whereas



NOTE: Positive paired scoring discordance indicates that site-based PANSS score was greater than site-independent score whereas negative discordance indicates the opposite.

Fig. 1. Distribution of Paired PANSS Scoring Discordance

NOTE: Positive paired scoring discordance indicates that site-based PANSS score was greater than site-independent score whereas negative discordance indicates the opposite.

site-based PANSS scores ≤ 80 generated statistically significant negative mean scoring discordance. The mean site-based score for PANSS scores ≥ 100 was 105.8 ± 5.1 whereas the mean paired site-independent score was 100.9 ± 9.5 ($t = 7.95$; $df = 172$; $p < 0.0001$). On the other hand, the mean site-based score for PANSS scores ≤ 80 was 67.5 ± 10.5 whereas the mean paired site-independent score was 74.0 ± 11.1 ($t = -6.69$; $df = 64$; $p < 0.0001$).

3.3. Comparison of site-based and site-independent PANSS treatment outcome

There were 148 subjects with paired site-based and site-independent PANSS data available from both the baseline and end of study visits. Subjects who lacked either baseline or endpoint/early termination PANSS review scores were not included in this comparative analysis. At the baseline visit, the mean PANSS total score for the 148 paired PANSS scores was 96.5 ± 8.5 (SD) for the site-based scores and 94.9 ± 9.3 for the site-independent scores ($t = 2.6$; $p = 0.01$).

As shown in Fig. 2, the 72 subjects assigned to the KarXT treatment group yielded significantly greater improvement from baseline to endpoint on the primary site-based PANSS total scores than the 76 subjects assigned to placebo ($F = 29.3$; $p < 0.0001$). Similarly, the blinded site-independent PANSS total scores derived from listening to and scoring the recorded site-based PANSS interviews also yielded significantly greater improvement favoring KarXT over placebo ($F = 12.5$; $p < 0.001$).

As shown in Table 1, both the site-based and site-independent PANSS scores achieved statistical significance favoring KarXT assigned subjects over placebo on the 4 key positive symptom items ($p = 0.0001$ and $p < 0.01$ respectively). Further, the KarXT assigned subjects achieved significantly greater improvement than placebo-assigned subjects on both the positive and negative symptom sub-scales based upon the primary site-based PANSS scores ($p < 0.0001$ and $p = 0.0005$ respectively). The site-independent PANSS positive score sub-scale was significant for

Table 1 Site-based and site-independent PANSS treatment outcome.

	Score change \pm SD (baseline to week 5/early termination)		
	KarXT	Placebo	
n	72	76	
PANSS total score			
site-based	-18.1 \pm 15.2	-5.3 \pm 13.6	F = 29.3; p < 0.0001
site-independent	-15.1 \pm 11.2	-8.1 \pm 12.7	F = 12.5; p < 0.001
4 key items^a			
site-based	-4.0 \pm 3.5	-1.8 \pm 3.2	F = 15.4; p = 0.0001
site-independent	-3.7 \pm 3.3	-2.4 \pm 2.8	F = 6.9; p < 0.01
PANSS positive sub-scale			
site-based	-5.4 \pm 4.7	-2.0 \pm 4.7	F = 20.1; p < 0.0001
site-independent	-4.8 \pm 4.4	-2.9 \pm 4.2	F = 6.8; p = 0.01
PANSS negative sub-scale			
site-based	-3.2 \pm 4.4	-0.7 \pm 4.1	F = 12.5; p = 0.0005
site-independent	-2.4 \pm 4.4	-1.7 \pm 3.7	F = 1.2; p = ns

^a 4 key items are derived from the PANSS positive syndrome scale: delusions, conceptual disorganization, hallucinatory behavior, and suspiciousness.

KarXT over placebo ($p = 0.01$), but not for the negative symptom sub-scale.

We used $\geq 30\%$ PANSS total score improvement from baseline to week 5/early termination as the criterion for treatment response. Using this criterion in this sample of 148 paired ratings, there were 30 KarXT responders (41.6%) and 8 placebo responders (10.5%) based upon the site-based PANSS scores ($\chi^2 = 18.8$; $df = 1$; $p < 0.0001$). There were 22 KarXT responders (30.6%) and 12 placebo responders (15.8%) based upon the site-independent PANSS scores ($\chi^2 = 4.6$; $df = 1$; $p = 0.03$). The site-independent raters matched 25 of the 38 site-based response scores of $\geq 30\%$ total PANSS score improvement (65.8% sensitivity) and 101 of the 110 non-response scores (91.8% specificity). The overall predictive value was 85.1% for matching the site-based response/non-response outcomes.

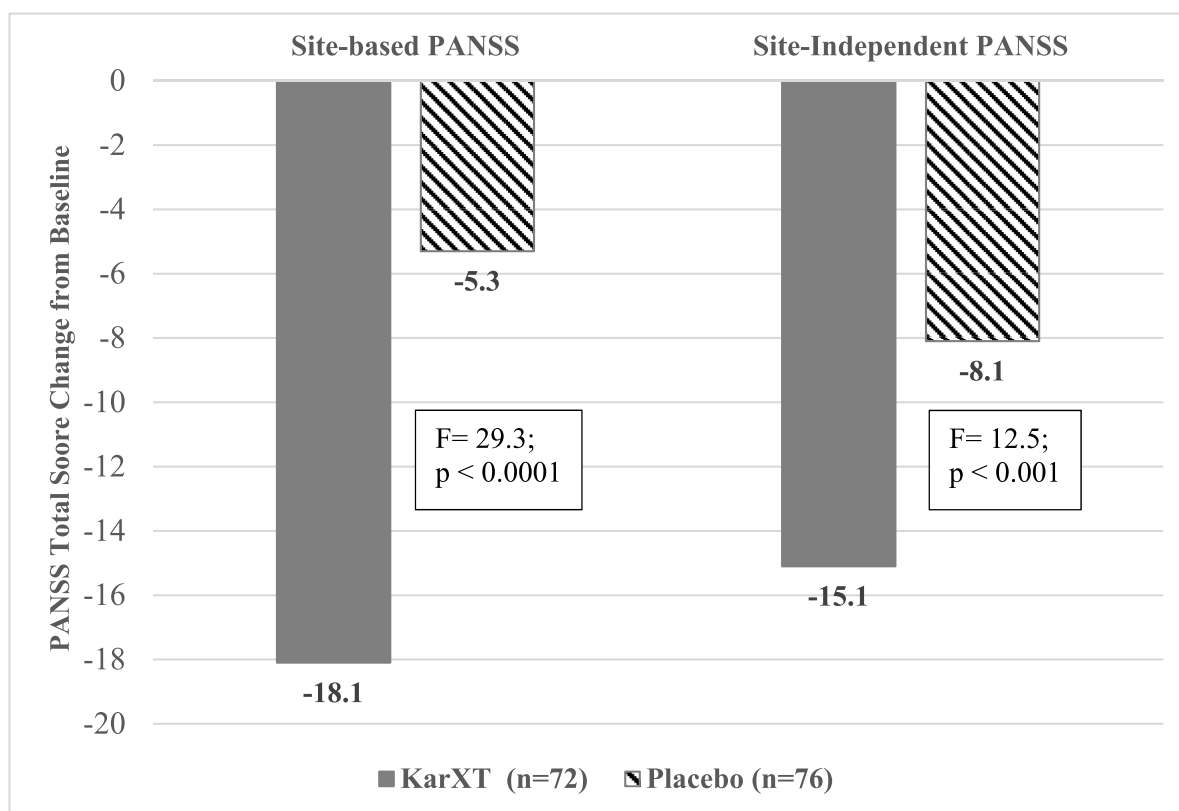


Fig. 2. Mean change from baseline of Total PANSS score.

3.4. Treatment emergent adverse events

The blinded, site-independent scores allowed us to examine the possibility of inadvertent functional unblinding by the site-based raters. A full description of TEAEs that occurred in this trial has been reported elsewhere (Brannan et al., 2021).

TEAEs were reported by 53.9% of subjects assigned to KarXT and 43.3% of subjects assigned to placebo in the safety population (Brannan et al., 2021). The TEAEs were primarily gastrointestinal symptoms (nausea, constipation, dry mouth, dyspepsia) and were noted more frequently in the subjects assigned to KarXT than the subjects receiving placebo. Only one subject had a serious TEAE. Most of the TEAEs were transient, mild to moderate symptoms that did not result in study withdrawal. Only 2 subjects assigned to placebo (2.2% of the safety population) and 2 subjects assigned to KarXT (2.2%) discontinued the study because of TEAEs.

4. Discussion

Drug developers rely on ratings accuracy and precision in clinical trials to properly assess drug efficacy and safety. In the present study, we examined ratings precision via “paired” ratings using site-based raters who administered the primary PANSS interviews and blinded site-independent raters who listened to and scored these audio-digital recorded interviews. A high correlation was demonstrated between site-based PANSS raters and site-independent raters on 561 recorded interviews across all visits ($r = 0.775$; $p < 0.0001$) for paired total PANSS score comparisons). This quality assurance monitoring strategy identified a few rater “outliers” for remediation during the study that may have positively impacted ratings precision for the remainder of the trial.

There were 148 paired PANSS ratings at the baseline and endpoint/early termination visits available for a comparative analysis of treatment outcomes. In this sample, subjects assigned to Kar-XT ($n = 72$) showed significant improvement from baseline to endpoint compared to subjects assigned to placebo ($n = 76$) as assessed by both the site-based PANSS total scores ($p < 0.0001$) and site-independent PANSS total scores ($p < 0.001$). These findings were supported by similar comparisons of the positive symptom sub-scale and 4 key positive items (Table 1).

TEAEs that occur during clinical trials have the potential to functionally unblind raters and study participants and impact ratings. More than 50% of KarXT assigned subjects and more than 40% of placebo-assigned subjects experienced at least one TEAE at some point in this study. Most reported TEAEs were mild and transient. The site-independent raters had no direct visual observation of the subject, had only the audio recorded PANSS interview and digital notes that addressed specific interview questions and were blinded to any adverse event information that might inadvertently have contributed to functional unblinding. The affirmation of the significant KarXT findings of the primary site-based PANSS ratings by the site-independent ratings mitigates possible concern about functional unblinding of site-based raters in this study.

The site-independent scores replicated the site-based PANSS findings but yielded less treatment effect than the site-based ratings in this very positive study. For instance, the KarXT treatment response rate was 41.6% as scored by the site-based raters in this sample of 148 paired ratings in contrast to 30.6% by the site-independent raters. This difference may reflect the inherent difficulty of obtaining remote assessments for psychiatric symptoms and particularly for negative symptoms in schizophrenia. Further, the slightly reduced site-independent signal detection may also be related to the bi-directional effect of symptom severity on scoring discordance. Site-independent raters tend to under-score higher severity subjects relative to site-based raters and overscore lower severity subjects. Thus, the site-based PANSS total scores that were ≥ 100 revealed significantly greater positive mean paired scoring discordance (site-based PANSS scores $>$ site-independent scores)

whereas lower site-based PANSS scores (≤ 80) had significantly greater negative mean scoring discordance. As a result of this bi-directional scoring effect, site-independent ratings are often lower than site-based PANSS scores at baseline (when symptom severity is highest) and slightly higher at the endpoint amongst responders. The narrower baseline to endpoint range of site-independent PANSS scores can reduce the amount a score change from baseline compared to the paired site-based scores. Consequently, the sensitivity of the site-independent scores to match the site-based treatment response scores (65.8%) was less robust than the specificity to match non-response scores (91.8%) because fewer site-independent change scores achieved the $\geq 30\%$ total PANSS improvement threshold for response between the baseline and endpoint. Nonetheless, the site-independent ratings achieved an overall predictive value of 85.1% for matching the site-based response/non-response outcomes.

The bi-directional effect of higher and lower symptom severity on scoring discordance noted in this study is consistent with an analysis of MADRS paired ratings conducted across 5 Major Depressive disorder (MDD) studies (Targum and Catania, 2019). Recently, we reported similar bi-directional effects of paired ratings for the Brief Psychiatric Rating Scale in other schizophrenia studies unrelated to the current report (Targum et al., 2021). Additional surveillance studies that examine the directional effect of paired ratings are needed in other clinical trials to explore the possible universality of this finding.

This was the first time we have been able to examine a sufficient number of paired baseline and endpoint scores in a schizophrenia study. Although site-based PANSS interviews were recorded and submitted for review at every visit, only a small number of interim visits between the baseline and endpoint visits were actually scored by site-independent raters in this study. Consequently, we could not compare paired scores at weeks 2 and 4 in this analysis. Therefore, this analysis is limited in that we could not examine serial change of paired independent scores from visit to visit.

In this study, we examined the utility of audio-digital recordings of site-based PANSS ratings for quality assurance purposes and for replication of site-based findings. Previous studies of audio-recorded MADRS interviews in MDD and BPRS in schizophrenia have demonstrated that site-independent ratings can replicate site-based treatment outcome ratings (Targum et al., 2014, 2019, 2021). A limitation of the audio-digital recording surveillance method is that the site-independent raters cannot have direct visual observation of the subject. Consequently, gestures, body posture, movement, and interactions with the interviewer cannot be adequately assessed. Although this limitation does not alter the primary QA function of the surveillance strategy, it was necessary to use the site-based score for 4 of the 30 PANSS items in the current study.

We used video recordings in a recently completed schizophrenia study unrelated to the current report that scored all 30 PANSS items; the ICC between the site-based and video-reviewed site-independent PANSS scores in that different study was $r = 0.839$ and consistent with the audio-digital method used in the current study (Targum, personal communication).

The high correlation between the site-based and site-independent PANSS ratings as well as the high predictive value demonstrated by the paired, blinded site-independent PANSS ratings in the EMERGENT-1 trial assures that the site-based PANSS scores are reliable and mitigates any concern about possible functional unblinding of site-based raters due to TEAEs at clinical trial sites. This method of QA assessment has now been shown to have utility in both MDD and schizophrenia studies and may have additional value in studies where functional unblinding is a concern.

Declaration of interests

Dr. Targum is an employee of Signant Health and has received vendor grants or consulted with Acadia Pharmaceuticals Inc., Alkermes

Inc., BioXcel Therapeutics Inc., Denovo Biopharma, EMA Wellness LLC, Epiodyne, Frequency Therapeutics, Functional Neuromodulation, Intra Cellular Therapies, Johnson and Johnson PRD, Karuna Therapeutics, Merck Inc., Methylation Sciences Inc., Navitor Pharmaceuticals Inc., and Sunovion Inc. during the past 3 years. Dr. Murphy is an employee of Signant Health. Dr. Breier is a Professor of Psychiatry at the Indiana University School of Medicine and is chief clinical advisor to Karuna Therapeutics and Dr. Brannan is an employee of Karuna Therapeutics.

Author statement

This manuscript represents original research that was designed and conducted by the co-authors of the report. The paper has not been submitted for publication elsewhere. All authors contributed to the analysis and writing of the manuscript as noted below and all authors have approved its submission in its current version.

Funding source

Funding for this analysis was provided by Karuna Therapeutics and Signant Health.

Acknowledgements

Sarah Kavanagh MPH contributed statistical review. Peter Weiden MD and Angel Angelov MD reviewed the manuscript.

References

- American Psychiatric Association, DSM5, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. American Psychiatric Association, Washington DC.
- Berendsen, S., Van, H.L., Verdegaal, L.M.A., van Tricht, M.J., Blankers, M., de Haan, L., 2020. Burying our heads in the sand: the neglected importance of reporting interrater reliability in antipsychotic medication trials. *Schizophr. Bull.* 46 (5), 1027–1029.
- Bodick, N.C., Offen, W.W., Levey, A.I., Cutler, N.R., Gauthier, S.G., Satlin, A., Shannon, H.E., Tollefson, G.D., Rasmussen, K., Bymaster, F.P., Hurley, D.J., Potter, W.Z., Paul, S.M., 1997. Effects of xanomeline, a selective muscarinic receptor agonist, on cognitive function and behavioral symptoms in Alzheimer disease. *Arch. Neurol.* 54 (4), 465–473.
- Brannan, S.K., Miller, A.C., Paul, S.M., Breier, A., 2018. Xanomeline plus trospium: a novel strategy to enhance pro-muscarinic efficacy and mitigate peripheral side effects. ACNP 57th Annual Meeting: poster Session I. *Neuropsychopharmacology* 43, 77–227.
- Brannan, S.K., Sawchak, S., Miller, A.C., Lieberman, J.A., Paul, S.M., Breier, A., 2021. Muscarinic cholinergic receptor agonist and peripheral antagonist for schizophrenia. *N. Engl. J. Med.* 384, 717–726.
- Correll, C.U., Kishimoto, T., Nielsen, J., Kane, J.M., 2011. Quantifying clinical relevance in the treatment of schizophrenia. *Clin. Therapeut.* 33 (12), B16–B39.
- Haro, J.M., Kamath, S.A., Ochoa, S., Novick, D., Rele, K., Fargas, A., et al., 2003. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatr. Scand. Suppl.* 16–23.
- Kavoussi, R., Miller, A.C., Brannan, S.K., Breier, A., 2017. Xanomeline plus trospium: a novel strategy to enhance pro-muscarinic efficacy and mitigate peripheral side effects. May 29–June 2, 2017. In: Poster Presented at American Society of Clinical Psychopharmacology. Miami Beach, FL.
- Kay, S.R., Opler, L.A., Fiszbein, A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276.
- Leucht, S., Davis, J.M., Engel, R.R., Kane, J.M., Wagenpfeil, S., 2007. Defining 'response' in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology* 32, 1903–1910.
- Leucht, S., Kissling, W., Davis, J.M., 2010. The PANSS should be rescaled. *Schizophr. Bull.* 36 (3), 461–462.
- Messer, W.S., 2002. The utility of muscarinic agonists in the treatment of Alzheimer's disease. *J. Mol. Neurosci.* 19 (1–2), 187–193.
- Muller, M.J., Szegedi, A., 2002. Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials. *J. Clin. Psychopharmacol.* 22, 318–325.
- Obermeier, M., Mayr, A., Schennach-Wolff, R., Seemüller, F., Möller, H.J., Riedel, M., 2010. Should the PANSS be rescaled? *Schizophr. Bull.* 36 (3), 455–460.
- Pak, R.W., Petrou, S.P., Staskin, D.R., 2003. Trospium chloride: a quaternary amine with unique pharmacologic properties. *Curr. Urol. Rep.* 4, 436–440.
- Posner, K., Brown, G.K., Stanley, B., Brent, D.A., Yershova, K.V., Oquendo, M.A., Currier, G.W., Melvin, G.A., Greenhill, L., Shen, S., Mann, J.J., 2011. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am. J. Psychiatr.* 168, 1266–1277.
- SAS Institute Inc, 2013. SAS/ACCESS® 9.4 Interface to ADABAS: Reference. SAS Institute Inc, Cary, North Carolina.
- Shannon, H.E., Bymaster, F.P., Calligaro, D.O., et al., 1994. Xanomeline: a novel muscarinic receptor agonist with functional selectivity for M1 receptors. *J. Pharmacol. Exp. Therapeut.* 269, 271–281.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., et al., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatr.* 59 (Suppl. 20), 22–33 quiz 4–57.
- Shekhar, A., Potter, W.Z., Lightfoot, J., et al., 2008. Selective muscarinic receptor agonist xanomeline as a novel treatment approach for schizophrenia. *Am. J. Psychiatr.* 165 (8), 1033–1039.
- Targum, S.D., Catania, C., 2019. Audio-digital recordings for surveillance in clinical trials of Major Depressive Disorder. *Contempor. Clin. Trials Commun.* <https://doi.org/10.1016/j.conctc.2019.100317>.
- Targum, S.D., Pendergrass, J.C., Toner, C., Asgharneshad, M., Burch, D.J., 2014. Audio-digital recordings used for independent confirmation of site-based MADRS interview scores. *Eur. Neuropsychopharmacol.* 24, 1760–1766.
- Targum, S.D., Pendergrass, J.C., Toner, C., Zumpano, L., Rauh, P., DeMartinis, N., 2015. Impact of interview length on ratings reliability in a schizophrenia trial. *Eur. Neuropsychopharmacol.* 25 (3), 312–318.
- Targum, S.D., Daly, E., Fedgchin, M., Cooper, K., Singh, J., 2019. Comparability of blinded remote and site-based assessments of response to esketamine or placebo nasal spray in patients with treatment resistant depression. *J. Psychiatr. Res.* 111, 68–73.
- Targum, S.D., Pendergrass, J.C., Murphy, C., 2021. Audio-digital recordings to assess ratings reliability in clinical trials of schizophrenia. *Schizophr. Res.* 232, 54–60.