Journal of Schizophrenia Research

Open Access

Austin Publishing Group

Review Article

Challenges and Putative Approaches to Improving Signal Detection in Schizophrenia Clinical Trials

Daniel DG[1*]; Busner J[2]; Kott A[3]

[1]Signant Health & George Washington University, USA

[2]Signant Health & Virginia Commonwealth University, USA

[3]Signant Health, Prague, Czech Republic

***Corresponding author: David Daniel**

Signant Health & George Washington University, 1071 Cedrus Lane, McLean, Virginia 22102, USA.

Tel: 703-638-2500

Email: david.daniel@signanthealth.com; dgdanielmd@gmail.com

Received: January 12, 2024

Accepted: February 12, 2024

Published: February 19, 2024

Abstract

Instruments for measuring symptom change in schizophrenia clinical trials are relatively complex and subjective compared to other CNS and non-CNS therapeutic areas, creating numerous challenges to detection of potential placebo-drug differences. To facilitate drug signal detection a plethora of interventions have been employed to putatively optimize selection, calibration and monitoring of raters in schizophrenia clinical trials and to control placebo response. Published literature describing and addressing the potential effectiveness of these methodologies is fragmented and relatively sparse. We describe the current and developing methodologies for optimizing data quality in schizophrenia clinical trials and discuss evidence bearing on their effectiveness. Awareness of these methodologies, their objectives and their limitations is important in planning and evaluating schizophrenia clinical trials.

Keywords: Clinical trials; Schizophrenia; Data quality; Rater training; Data quality monitoring

Introduction

Multiple factors challenge signal detection in schizophrenia clinical trials, including insufficient understanding of the biological mechanisms underlying schizophrenic psychopathology, inadequacy of trial designs, challenges in patient selection, and marginal sufficiency of efficacy endpoints [1,2]. In recent years, placebo response has increased while drug response has remained stable in acute schizophrenia clinical trials and there have been recent, unexpected phase 3 acute schizophrenia trial failures following robust phase 2 success [1]. In phase 3 clinical trials with stable schizophrenic patients with predominantly negative symptoms, robust placebo-drug separation has also been challenging and no pharmacological treatments have, to date, clearly demonstrated effectiveness [2,3].

Compared to other CNS and non-CNS therapeutic areas, rating scales utilized in in schizophrenia clinical trials, especially those used to assess negative symptoms, are relatively complex and subjective. This presents a plethora of challenges for the investigator, who is required to measure symptom severity with accuracy and precision while modulating expectation bias on the part of the patient and informant that might enhance placebo response. Schizophrenia clinical trial ratings calibration exercises typically address the Positive and Negative Syndrome Scale (PANSS) [4]. Reviews of recorded site interviews by independent reviewers suggest that raters tend to have more difficulty reliably rating PANSS items based on objective observations of behavior compared to PANSS items rated by verbal report [5]. Site raters had the lowest concordance with external reviewers when rating negative symptoms, especially blunted affect, poor rapport, and lack of spontaneity of conversation [6]. In a survey of 39 raters participating in an industry sponsored clinical trial, fewer than 11% evaluated the PANSS negative symptom or Negative Symptom Assessment (NSA-16) anchor points as “Very clear” [7,8].

Factors modulating successful selection and calibration of raters and their performance rating subjects once the study is underway are poorly understood [9]. Phase 3 trials may be vulnerable to failure after successful phase 2 trials due to expectation bias and greater challenges calibrating a larger universe of sites, languages, and cultures.

Recently, recruiting periods for numerous schizophrenia clinical trials have been extended due to insufficient clinical trials sites and raters in the wake of geopolitical conflict in Eastern Europe and the COVID pandemic. With the field experiencing shortages of experienced, high quality schizophrenia clinical trial raters to service ongoing and planned studies, the need for effective methodologies for selection of raters, calibration of symptom measurement and effective endpoint data quality monitoring has taken on increased urgency. Shown in Table 1 are comprehensive procedures for establishing and maintaining accurate, calibrated ratings in schizophrenia clinical trials that have been widely adopted by industry. The burden to raters

Citation: Daniel DG, Busner J, Kott A. Challenges and Putative Approaches to Improving Signal Detection in Schizophrenia Clinical Trials. J Schizophr Res. 2024; 10(1): 1047.

Table 1: Examples of Procedures Putatively Optimizing Endpoint Data Quality in Schizophrenia Clinical Trials.

Site and rater selection based on previous performance
Pre-study calibration of diagnostic assessment, interview technique and symptom severity measurement
Placebo response modulation training for the research site, patient and informant
Placebo response mitigation scripts incorporated into ratings procedures
Standardized, supportive psychotherapy
Enhanced rating instructions and consistency checks embedded in eCOA
Recording and independent expert review of rating interviews and scoring
Blinded analytic review of endpoint data to detect aberrant rating patterns
Rapid remediation of rating and interview errors
Site enrollment continually tied to assessment of data quality
Poorly performing sites remediated or closed
Use of inpatient setting in trials of acutely exacerbated patients to reduce measurement noise associated with medication non-compliance, drug abuse and environmental stress

and expense to clients of these procedures are considerable. Industry-wide attempts to share fragmented rater training and performance data to reduce redundancy of training and quality assurance procedures have, unfortunately, met with limited success. In 2014, the CNS Summit Rater Training and Certification Committee convened a panel at the Summit's annual meeting entitled "Has it been worth it? A Critical Appraisal." The panel published a consensus statement on recommended training and monitoring procedures [9]. A decade later, with respect to these procedures, raters and clients continue to ask, "Is it worth it?". In this paper the authors discuss observations bearing on the question of "is it worth it?" presented at scientific meetings and in published literature over the last decade and a half.

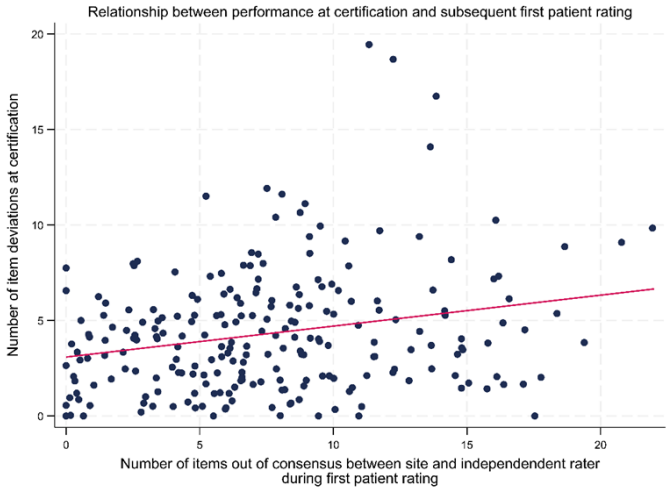

Figure 1: Rater Training Performance Predicts Quality of Subject Ratings [11].

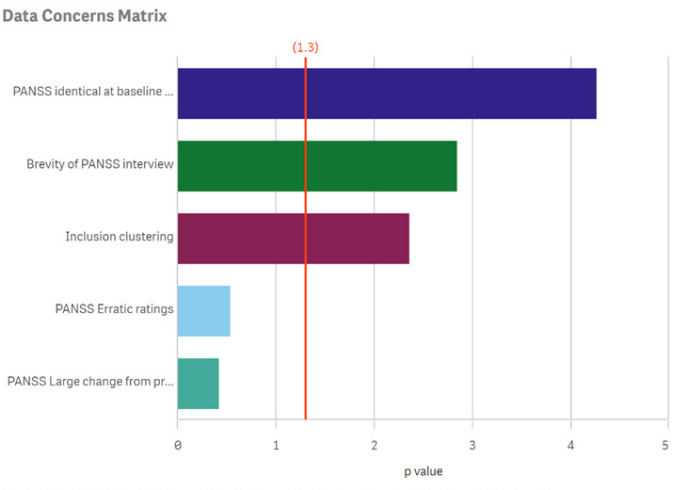

Figure 2: Examples of Anomalous Rating Patterns at a Site of Concern.

Precision in Measurement Among Investigators Impacts Sample Size Requirements and is Readily Achievable

The impact of calibration and reliability of ratings on sample size, statistical power, and the ability to detect placebo-drug differences in clinical trials is well documented [10]. Empirically demonstratable benefits from calibration of raters include an increased level of confidence in trial results and cost and time savings from smaller sample sizes.

Rater training typically includes a slide review addressing best practices for administering and scoring each rating scale followed by a group rating calibration exercise of a videotaped interview or an on-stage patient or actor. To be approved to rate in the study, investigators are required to rate a full-length interview of the primary efficacy scale consistent with panel-based gold standards and group norms [4]. The training can be divided among asynchronous and real time on-line or in-person components. There is little credible empirical evidence that in-person training is superior to virtual training to achieve scoring calibration when controlled for experience and credential levels. However, many sponsors and investigators prefer the in-person experience. As shown in Figure 1, a rater's performance in the certification process to rate the PANSS appears to be modestly but statistically significantly predictive of performance rating patients at the site [11].

Lack of consistency in interviewing practices may alter the patient's responses and obscure any potential drug signal. Interview of a live actor portraying a subject may be employed to assess and calibrate raters' interview practices [12]. Sufficient probing to distinguish among the anchor points of lengthy rating scales, objectivity, and efforts to neutralize expectation bias and thus reduce placebo response should be evaluated [12,13]. Semi-structured interviews for schizophrenia rating scales such as the Structured Clinical Interview for the PANSS (SCI-PANSS) and Negative Symptom Assessment Scale (NSA-16) Manual have been employed to support thoroughness and consistency of interview technique [8,14,15]. Van Knorring et al (1995) reported a modest increase in inter-rater reliability when the SCI-PANSS was used compared to the PANSS alone [14]. However, the SCI-PANSS does not sufficiently query the frequency, severity, and impact of symptoms to facilitate distinguishing among the PANSS anchor points and fails in re-establishing the time frame leading to possible contamination of ratings by symptoms not present in the required timeframe. The SCI-PANSS is designed to address verbal reports from the subject but does not address the informant or behavioral observations of the subject required to arrive at the score of numerous PANSS items. Elements of the SCI-PANSS script are inapplicable in some cultures.

Inexperienced raters should be cautioned to administer the script flexibly and never in a rote manner.

Training and standardization of interviewing procedures typically focus on directly assessing the patient. However, the basis of rating numerous PANSS questions includes the informant [16]. Not including the informant information, as sometimes done in clinical trials, appears to result in lower PANSS total scores and reduced changes in symptom severity over time [16]. Further, inconsistent use of informant information across visits may obscure the study signal.

Informants, like patients, may be subject to expectation bias that can impact placebo response. Thus, PANSS interview training should usefully focus on both the patient and informant. The Informant Questionnaire (IQ–PANSS) is sometimes utilized in schizophrenia clinical trials to assure informant information is systematically collected [16]. Once the study is underway, rating scale interviews of both the patient and informant may be recorded for external review of rating and interview quality.

Critical but sometimes ignored aspects of interview training are placebo response mitigation measures such as reduction of expectation bias and dissuasion of the natural tendency to guess treatment allocation. The former may be a particularly potent source of placebo response in phase 3 trials due to positive expectations from successful phase 2 trials. For optimal effect, placebo response mitigation training measures should directly address the rater, patient, informant, and everyone else at the site who has contact with the patient and informant. Cohen and colleagues (2021) observed that in subjects with psychotic and major depressive disorders, a participant-focused psychoeducational procedure, educating and subsequently reminding participants about key factors known to amplify placebo response, was associated with a systematic reduction in symptom reports and global subjective impressions of change over the study period [17].

Remote administration of existing schizophrenia rating scales by phone or audio-video technology was done sporadically out of necessity during the COVID-19 pandemic. Training should involve synchronization of remote administration of clinician-administered scales with comparison to in-person administration in the same subjects [18]. Audio-video assessment is preferred over audio alone because the basis of rating of many scale questions includes visual assessment.

Initial calibration of rating technique is feasible across linguistically and culturally diverse regions including North America, Eastern and Western Europe, Central and South America, South Africa, and Australia with overall kappas of 0.84 for the PANSS negative subscale and .89 for the NSA-16 [19]. However, following initial calibration, there is sparse evidence to inform the frequency, if any, that rater training should be repeated (commonly referred to as “refresher training”) to maintain calibration. In a retrospective analysis of rater performance in rating a videotaped PANSS or NSA-16 interview at mid-study, we noted similar levels of rater calibration compared to study initiation [20]. Without a comparison group it was not possible to determine whether the high rate of rater agreement seen at mid-study was related to the refresher training procedures vs. the experience of rating the scales during the study or both.

Data Quality Issues are Common Even Among Experienced, Trained Investigators

In a large sample of clinical trial PANSS ratings, Rabinowitz et al found that almost 40% of PANSS study visits had at least one inconsistency flag raised and 10% had two [21]. This mirrors our experience in which a wide variety of data anomalies are detected even among experienced, well vetted raters. Examples include logical inconsistencies of measurement of related constructs within and across scales, erratic scoring patterns, identical ratings from visit to visit, clustering of severity scores near entry criteria at screening and poor interview quality. The composition of specific data quality issues tends to vary across geographic regions [22,23]. The subjectivity of the rating instruments appears to present challenges to maintenance of ratings calibration even among the most skilled, seasoned raters.

Figure 2 illustrates how the prevalence of rating anomalies at an individual research site can be profiled in comparison to peer sites within the same clinical trial. In this example, the frequency of poor interview quality and the other quality indicators that cross the vertical red line are statistically significant outliers compared to the other clinical trial sites in the study. The identification of outlying sites provides an opportunity for constructive remediation of erroneous interview and rating practices or in extreme situations, limiting enrollment at the site.

Figure 3 illustrates how sites can be ranked comparatively based on composite data quality indicators. These rankings can inform which sites receive remediation as well as allocations of additional subjects. Moreover, the rankings can aid in site selection for future trials. In rater selection, clinical and scale experience requirements are usually rigorous. There is a relatively sparse body of literature consistent with the notion that experience, credentials and training are predictive of the quality of endpoint data produced by a rater once the trial is underway. For example, in a retrospective analysis of 957 raters intending to rate the PANSS in acute schizophrenia trials, we found that years of clinical trial experience was predictive of the number of deviations from an expert panel in rating a videotaped PANSS interview [24]. Doctorate level raters exhibited greater competency and less variability in conducting PANSS interviews compared to non-doctorate raters as evaluated by the Research Interview Assessment Scale (RISA), which assesses a broad range of interview behaviors [25]. In a sample of 30 subjects administered the Hamilton Depression Scale (HAM-D), Kobak and colleagues (2009) observed that raters who were both experienced and calibrated had the highest interrater reliability (intraclass correlation [ICC]; r=0.93) followed by inexperienced raters (r=0.77) followed by experienced but uncalibrated raters (r=0.55) [26]. The results are consistent with the need for both calibration and experience.

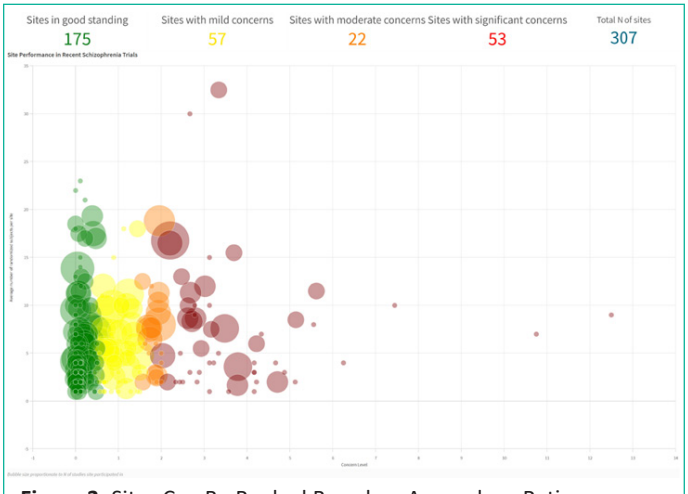

Figure 3: Sites Can Be Ranked Based on Anomalous Rating Patterns.

Rater Performance Varies by Geographic Region

Multiple measures of rater performance vary by geographic region. North American raters scored modestly worse than non-North American raters on the RISA [25,27]. In schizophrenia trials, non-doctorate level raters are more commonly relied upon in the United States compared to the rest of the world, particularly Europe [27]. However, as evaluated by the Rater Quality Questionnaire (RQQ), which focuses specifically on the quality of information collected during the interview and on adherence to rating scale rules, North American Raters scored as well or better than their colleagues in other parts of the world [23,28]. Rating anomalies, such as discordance between the PANSS and Clinical Global Impression (CGI) scales, also vary in frequency by geographic region, with comparatively fewer errors in eastern Europe [29].

Despite standardized training, modest but statistically significant differences are observable by region in the severity of negative symptoms measured by the PANSS and NSA-16 scores at study baseline [30,31]. Although insufficient global calibration may be partially responsible for regional differences, it is likely that cultural impact on expression of schizophrenia, which is well documented, is also a factor [31].

Research subject recruitment rates are impacted by trial type and geography. In trials involving acutely decompensated schizophrenic patients North American investigators recruited at a significantly higher rate than Asian and Eastern European investigators [32]. In clinical trials involving stable schizophrenic patients with predominantly negative symptoms, both Eastern Europe and South America had significantly higher recruitment rates compared to Asia and North America [32].

In our experience, despite overall regional differences in quality metrics, individual sites within regions vary markedly in experience and data quality. In a post-study survey, Loebel et al (2010) noted several schizophrenia clinical site characteristics influencing the likelihood of detecting a placebo vs. drug difference, including the source of referral of patients, proportion of research experienced patients, proportion of pharma sponsored revenue from industry sponsored studies, and the extent to which the principal investigator values placebo response mitigation practices [33].

Many Data Quality Issues are Associated with Remediable Rating Practices

As shown in Figure 4, data quality issues detected early in schizophrenia clinical trials are highly predictive of recurrence after randomization [34]. Data quality aberrations, including increased and decreased variability, may impact placebo response and drug response differentially with a detrimental impact on drug-placebo separation [35]. For example, high within subject visit-to-visit variability, including erratic changes, has been shown to be associated with increased placebo response and diminished signal detection in both acute schizophrenia and prominent negative symptom clinical trials [35,36]. Sites with a high frequency of erratic ratings are easily identified in blinded data and warrant scrutiny for potential frequent rater change, inconsistency of interviewing and rating technique, subject selection anomalies, medication non-compliance and unstable ward environments [35-38]. High within subject variance appears to be associated with multiple other data quality issues, including PANSS logical inconsistencies and CGI-PANSS inconsistencies [37,38]. Rater change, a modifiable site behavior, is associated with large changes across visits in the total PANSS score and increase in within subject variability, but this increase is not seen consistently across all 5 PANSS factors [35,37]. Variation in the time of day of assessment is associated with increased same patient visit to visit variability in PANSS scores [39]. Short PANSS interviews (eg, less than 20 minutes) are associated with a variety of data quality issues compared to more standard interviews. [40].

Identical scoring of all 30 items of the PANSS across visits, especially in the context of rater change, is felt to raise questions about whether the study PANSS interviews and scoring procedures for those visits were conducted independently of each other [41].

Logical inconsistencies in scoring items within the PANSS can be deterred by careful attention to rating instructions and anchor point descriptions as well as software programming within the electronic clinical outcome assessment (eCOA) which advises the rater of the potential incompatibility of PANSS item scores. Research sites with outlying numbers of PANSS logical inconsistencies are at risk for a higher response to placebo than non-outlying sites [42].

Inexplicable scoring discrepancies between the change from baseline in the CGI-S and total PANSS scores may be driven by non-communicating raters scoring the PANSS and CGI-S [43]. On the other hand, when a few PANSS items are exerting disproportionate influence on the patient's clinical condition, for example, scoring discrepancies between the CGI-S and total PANSS score may be accurate. Common causes of discrepancies between the CGI-I and change from baseline in the CGI-S and PANSS scores are referencing the CGI-I to a visit other than baseline and scoring the CGI-I out of order [44,45]. The CGI-I should be informed by and scored after the other efficacy scales. Administering scales in the incorrect order and gross scoring incompatibility errors within and across scales can be deterred by educational procedures and eCOA platforms that require correct scale administration order, per the protocol, or flag major discrepancies before data submission [45]. In the case of the latter, raters are given the option to respond to the flag, but are not required to do so.

Centralized analysis of blinded data for aberrant patient selection patterns and rating anomalies can be paired with audio/video recording of subject interviews to cost-effectively identify sites at risk for signal degradation [46,47]. With close monitoring of ratings quality, feedback and remediation, the frequency of errors in rating the PANSS and CGI appears to fall statistically

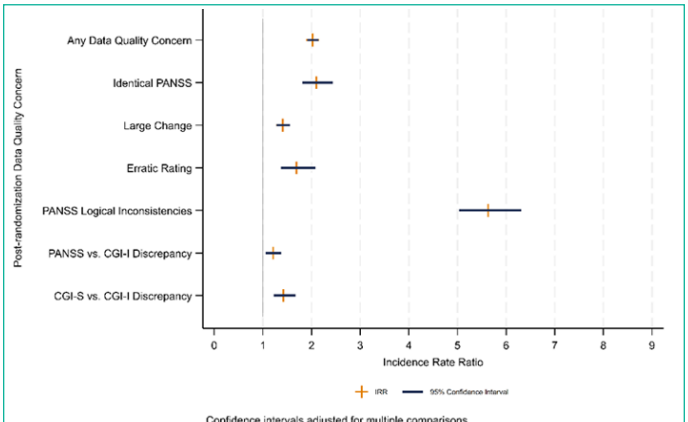

Figure 4: Effect of presence of any data quality concern before randomization on the incidence post-randomization data quality concerns N = 10,056 Subjects.

significantly over a six-month period, consistent with improvement in rating technique [48]. Audio/video recording coupled with external expert review of site PANSS interviews appears to reduce identical ratings of 30/30 PANSS items across consecutive visits (a putative measure of non-independent PANSS assessments) by over 50% [49]. Combining eCOA with audio/video recording further reduces the frequency of scoring errors [50]. Audio recording is often viewed as less intrusive and more conducive to patient confidentiality compared to video recording. However, rating scales such as the PANSS and NSA-16 have significant components that are evaluated visually. Thus, audio-video recording provides a more thorough assessment than audio alone and higher agreement between site and external raters [51]. Nevertheless, blinded, site-independent PANSS ratings derived from listening to and scoring audio recorded site-based interviews have high overall predictive value for matching site-based ratings [52]. Audio recorded site-based interviews may have further utility in avoiding detection of "functional" treatment emergent adverse events that may bias ratings [52,53].

With surveillance of ratings and ongoing feedback to investigators, the quality of interview data and proficiency of ratings were judged to be adequate or better by external reviews in the large majority of cases [28]. In post-hoc analyses of schizophrenia clinical trials comparable in design, enhanced data quality assurance methods such as those shown in Table 1 appear to be associated with fewer clinically meaningful data quality concerns [54,55].

Patient reported outcomes, especially ecological momentary assessment of subject activity, are increasingly incorporated into outpatient schizophrenia clinical trials [56]. Quality concerns also occur at a high frequency in Patient Reported Outcomes (PRO) data. Concerning patterns can be easily detected in blinded electronic (ePRO) data either by visual inspection or programmed quality indicator alerts. Examples include implausible values, repetitive responses, unexpected variability and unusual administration times and time stamps [57].

Barriers to Diversity in Clinical Trial Recruitment are Numerous but can be Addressed by Multiple Means

Racial and ethnic disparities in schizophrenia and other clinical trial participation are well documented [58]. A recent survey of clinical trialists noted numerous obstacles to clinical trial recruitment of Underrepresented and Marginalized Groups (UMB), including cultural beliefs, linguistic barriers, perceived lack of interest and lack of information [59]. Strategies proposed to improve recruitment included engagement with community leaders, targeted advertising, utilizing databases, and social media campaigns [59].

Machine Learning Can Identify at Risk Sites and Raters

Machine learning offers the opportunity to enhance proactive identification of raters and sites at risk of developing data quality concerns for early remediation or limitations on enrollment. The recent advances in machine learning offer an opportunity to prospectively identify raters and sites at risk of developing future data quality concerns throughout the study. It is however imperative that only highly accurate and clinically relevant models providing actionable predictions are considered as the application of inaccurate or irrelevant models may result in data quality deterioration [60].

We have demonstrated successful implementation and 12-month stability of two complex machine learning pipelines predicting high variability and within PANSS discrepancies [61]. Machine learning also offers the opportunity to seamlessly assess subjects' suitability for a clinical trial or monitor rater performance and other, currently unforeseen, applications are likely to emerge as the methodologies further evolve.

Age Matters: Inclusion of Adolescent Participants in Schizophrenia Trials Warrants Specialized Training, Specialized measures, and Focused Attention on Data Quality

Along with many welcome pediatric regulatory initiatives are those incentivizing and at times mandating pharmaceutical sponsors to include patients aged 13-17 in their schizophrenia trials [62].

Schizophrenia is less common in adolescents than in adults, and there is often diagnostic ambiguity in the presentation and/or reluctance on the part of practitioners to make a schizophrenia diagnosis even when the criteria are clearly met [63]. In addition to the difficulty of securing appropriately diagnosed patients, the relatively modest pool of investigators trained in child and adolescent psychiatry in the US, and even more so outside of the US, represents an additional challenge when designing and conducting clinical trials in adolescents with schizophrenia [64].

Another challenge comes from the measures themselves – such as the PANSS -- designed for adults but used ubiquitously as the primary efficacy measure in adolescent schizophrenia trials [65]. Conventions have emerged over the years for interviewing the parent/caregiver, as well as the patient, on each of the 30 PANSS items in adolescent trials. This is different from what is done in adults and adds another layer of complexity for investigators not experienced or skilled in working with this population.

In addition to the learnings relative to adult patients with schizophrenia, as discussed throughout this paper, are learnings unique to the adolescent population.

Diagnosis in Pediatric Trials: Following focused expert training on the symptomatic presentation and differential diagnostic considerations of the disorder, we recommend external expert review of diagnostic interviews and outside verification of the diagnostic eligibility of each selected participant.

Efficacy Assessment in Pediatric Schizophrenia Trials: As true for studies with adults, we recommend external review of PANSS interviews for interview adequacy and scoring appropriateness. Regulators often allow an allotted number of adolescents into adult trials, and it is not uncommon for sponsors to allow adult-trained investigators to enroll adolescents into their ongoing schizophrenia trials. Investigators who have worked in adult studies may not adhere to the special PANSS conventions for adolescents and are often not versed in probing/following up/scoring PANSS items in the adolescent age group. Our group has shown there to be high variability amongst PANSS items when raters attempt to score standardized adolescent patients with schizophrenia using the PANSS [66,67].

Recent Advances in Pediatric Schizophrenia Trials

In an effort to improve signal detection and reduce burden, much research has been devoted to shortening the PANSS for specific use in the 13–17-year-old population; a 10 item psychometrically derived version has been developed from a government funded trial of schizophrenic adolescents, and findings have now been replicated in 2 large independent industry

sponsored pivotal trials with schizophrenia adolescents [68-70].

In addition, a structured interview that assists raters in appropriately querying, probing, and scoring the 10 items is in the final stages of development, as is an eCOA version that will provide independent quality assurance metrics to help identify potential rating errors [71].

Conclusions

Returning to the question asked earlier, "Is it worth it?", we have presented a number of observations consistent with a qualified "yes". That is, there appears to be a limited, but measurable benefit to endpoint data quality from many of the rater centered procedures described. Moreover, certain putatively detrimental data quality indicators (e.g., erratic ratings) appear to be associated with increased placebo response and diminished placebo-drug separation. While these results are consistent with a beneficial effect of rigorous training and data monitoring, interpretation is limited by the post-hoc nature of the analyses and the often uncontrolled or inadequately controlled nature of the comparisons. Among salient future directions of research are how much training and data quality monitoring is enough; the extent to which high quality data and placebo-drug separation at a site are state vs. trait phenomena; and how accurately a site's pattern of quality indicators in blinded data predicts drug-placebo separation.

Author Statements

Funding

Signant Health

References

1. Leucht S, Chaimani A, Mavridis D, Leucht C, Huhn M, Helfer B, et al. Disconnection of drug-response and placebo-response in acute-phase antipsychotic drug trials on schizophrenia? Meta-regression analysis. Neuropsychopharmacology. 2019; 44: 1955-1966.

2. Marder SR, Davidson M, Zaragoza S, Kott A, Khan A, et al. Issues and Perspectives in Designing Clinical Trials for Negative Symptoms in Schizophrenia: Consensus Statements. Schizophrenia Bulletin Open. 2020; 1.

3. Fraguas D, Diaz-Caneja CM, Pina-Camacho L, Umbricht D, Arango C. Predictors of placebo response in pharmacological clinical trials of negative symptoms in schizophrenia:a meta-regression analysis. Schizophr Bull. 2019; 45: 57–68.

4. Opler MGA, Yavorsky C, Daniel DG. Positive and Negative Syndrome Scale (PANSS) Training: Challenges, Solutions, and Future Directions. Innov Clin Neurosci. 2017; 14: 77–81.

5. Daniel DG, Dries J. What PANSS Items Do Site Raters Have the Most Trouble Rating? In: Poster Presentation, 53rd Annual New Clinical Drug Development Unit Meeting. Hollywood, FL; May 28-31 2013.

6. Daniel DG, Kott A. Poster# M162 Which Negative Symptoms Do Raters Have the Most Trouble Rating. Schizophrenia Research. 2014; 153: 249.

7. Daniel DG, Velligan D, Greco N, Bartko JJ. Comparing measures of negative symptoms of schizophrenia in clinical trials: The Investigators' View. In: Poster presentation at 7th Annual International Society for Clinical trials Methodology (ISCTM) Scientific Meeting. Washington, DC. 2011; 21-23.

8. Axelrod BN, Goldman RS, Alphs LD. Validation of the 16-item Negative Symptom Assessment. J Psychiatr Res. 1993; 27: 253-8.

9. West MD, Daniel DG, Opler M, Wise-Rankovic A, Kalali A. CNS Summit Rater Training and Certification Committee Consensus recommendations on rater training and certification. 2014.

10. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. Biol Psychiatry. 2000; 47: 762-6.

11. Daniel D, Kott A. Do Rater Certification Procedures Identify Poor Raters? In: Poster presentation at Fall Meeting of the Annual International Society for CNS Clinical trials and Methodology (ISCTM) Scientific Meeting. Philadelphia, PA. 2013.

12. Kobak KA, Opler MGA, Engelhardt N. PANSS rater training using Internet and videoconference: Results from a pilot study. Schizophrenia Research. 2007; 92: 1-3.

13. Popp D, Kobak K, Detke M. P-640–- the Power of Expectation Bias. European Psychiatry. 2012; 27: 1–1.

14. Knorring L, Lindström E. Principal components and further possibilities with the PANSS. Acta Psychiatrica Scandinavica. 1995; 92: 5–10.

15. Alphs L, Velligan D, Daniel DG. The eCOA Negative Symptom Assessment-16 (NSA-16) Instruction Manual – Version 3.0. In: Annual International Society for Clinical trials Methodology (ISCTM) Scientific Meeting. Washington, DC, February, 16-18 2016.

16. Nielsen CM, Kølbæk P, Dines D, Opler M, Correll CU, Østergaard SD. Are informants required to obtain valid ratings on the Positive and Negative Syndrome Scale (PANSS)? Schizophrenia (Heidelberg, Germany). 2023; 9: 54.

17. Cohen EA, Hassman HH, Ereshefsky L, Walling DP, Grindell VM, Keefe RSE, et al. Placebo response mitigation with a participant-focused psychoeducational procedure: a randomized, single-blind, all placebo study in major depressive and psychotic disorders. Neuropsychopharmacology. 2021; 46: 844-850.

18. Daniel DG, Cohen AS, Velligan D, Harvey PD, Alphs L, et al. Remote Assessment of Negative Symptoms of Schizophrenia. Schizophrenia Bulletin Open. 2023; 4.

19. Daniel DG, Alphs L, Cazorla P, Bartko JJ, Panagides J. Training for assessment of negative symptoms of schizophrenia across languages and cultures: comparison of the NSA-16 with the PANSS Negative Subscale and Negative Symptom factor. Clin Schizophr Relat Psychoses. 2011; 5: 87-94.

20. Daniel DG, Bartko J, Allen M. Assessment of Rater Drift in CNS Clinical Trials. In: APA Annual Meeting. Washington, DC. 2008; 3-8.

21. Rabinowitz J, Schooler NR, Anderson AE, Ayearst LE, Daniel DG, Davidson MH, et al. Consistency checks to improve measurement with the Positive and Negative Syndrome Scale (PANSS. Schizophrenia Research. 2017; 190: 74–6.

22. Daniel D, Kott A. "p. 3. d. 088 Regional variation in data analytic findings in global schizophrenia clinical trials. European Neuropsychopharmacology. 2015; 25: 532.

23. Daniel DG, Kott A. How Do US Clinical Research Sites Compare with Rest of World in Interview and Ratings Quality? In: Poster presentation at 166th Annual Meeting of the American Psychiatric Association (NR6-18). San Francisco, CA. 2013

24. Daniel D, Wang X, Iacob A, EP, Kott A. Do Experience and Credentials Impact Clinical Trial Performance?. Innov Clin Neur. 2022; S18: 10–2.

25. Daniel D, Bartko J, Sartorius N, Vieta E. Educational level of raters impacts interview quality in international antipsychotic clinical trials. Schizophrenia Research. 2010; 2: 189.

26. Kobak KA, Brown B, Sharp I, Levy-Mack H, Wells K, Ockun F, et al.

Sources of unreliability in depression ratings. J Clin Psychopharmacol. 2009; 29: 82-5.

27. Daniel D, Bartko J, Sartorius N, Vieta E, Butler A, Moya G. P.2. e. 032 Patterns in European and rest of world use of doctorate level raters in bipolar and schizophrenia clinical trials. European Neuropsychopharmacology. 2009; 19: 466–7.

28. Daniel D, Kott A. "p. 3. c. 038 Initial experience with the Ratings Quality Questionnaire: a new tool for evaluating quality of clinical trials ratings. European Neuropsychopharmacology. 2012; 22: 338–9.

29. Daniel D, Kott A. Frequency and Regional Distribution of Errors Scoring the CGI Detected by Blinded Data Analytics. In: Poster presentation at the 11th Annual Meeting of international Society for CNS Trials and Methodology (ISCTM). Washington, DC. 2015.

30. Kott A, Wang X, Daniel DG. Regional Differences in NSA-16 Factor Scores at Study Entry: An Exploratory Analysis. In: Poster presentation at the Congress of the Schizophrenia International Research Society (SIRS) April. 2022; 6-10.

31. Khan A, Liharska L, Harvey PD, Atkins A, Ulshen D, Keefe RSE. Negative Symptom Dimensions of the Positive and Negative Syndrome Scale Across Geographical Regions: Implications for Social, Linguistic, and Cultural Consistency. Innov Clin Neurosci. 2017; 14: 30–40.

32. Daniel D, Kott A. Regional and Population Differences in Schizophrenia Clinical Trial Recruitment Rates. In: Poster presentation at the 18th Annual Meeting of the International Society for CNS Clinical Trials and Methodology (ISCTM). Washington, DC. 2022.

33. Loebel A, Cucchiaro J, Daniel D, Kalali A. Signal Detection in Clinical Trials: A Post-Study Survey of Schizophrenia Trial Sites. In: Poster presentation, International Society Clinical Trials Methodology, Autumn Conference. Baltimore, Maryland. https: // isctm.org/meeting-archives. 2010.

34. Kott A, Daniel D. Early Indicators of Poor Data quality in Schizophrenia Clinical Trials. In: 12th Annual Meeting of the International Society for Clinical Trials Methodology (ISCTM). Washington, DC. 2016.

35. Kott A, Brannan S, Wang X, Daniel D. The Impact of Aberrant Data Variability on Drug-Placebo Separation and Drug/Placebo Response in an Acute Schizophrenia Clinical Trial. Schizophr Bull Open. 2021; 2: sgab037.

36. Umbricht DS, Kott A, Daniel DG. The Effects of Erratic Ratings on Placebo Response and Signal Detection in the Roche Bitopertin Phase 3 Negative Symptom Studies—A Post Hoc Analysis. Schizophrenia Bulletin Open. 2020.

37. Kott A, Wang X, Daniel D. Exploring the Association of PANSS Rater Change with Extreme Within-Subject Variability in Schizophrenia Clinical Trials. In: ISCTM 15th Annual Meeting. Washington, DC; February 19-21 2019.

38. Kott A, Wang X, Sachs G, Daniel D. Within person variance as a quality metric–an exploratory analysis identifying outlier sites in schizophrenia clinical trials. European Neuropsychopharmacology. 27.

39. Daniel DG, Kott A. Does variation in the time of day of PANSS assessment effect symptom severity? Poster presentation. In: International Society for CNS Clinical Trials and Methodology (ISCTM). Washington, D.C., USA. 2018.

40. Kott A, Daniel DG. Association of PANSS interview duration with data quality – an exploratory analysis. In: Poster presentation International Society for CNS Clinical Trials and Methodology (ISCTM). Washington, DC, USA. 2018.

41. Kott A, Daniel D. Rater Change Associated with Identical Scoring of the PANSS as a Marker of Poor Data Quality. Innovations in Clinical Neuroscience. 2015; 12: 1–20.

42. Kott A, Lee J, Forbes A, Pfister S, Ouyang J, Wang X, et al. Logical Inconsistencies Among PANSS Items are Associated with Greater Placebo Response in Acute Schizophrenia Trials. In: 12th Annual International Society for CNS Clinical Trials and Methodology (ISCTM) Fall Conference. Philadelphia, Pennsylvania; September 26-27 2016.

43. Kott A, Daniel D. Discrepancies between CGI-S Score and PANSS Level Scores-an Exploratory Analysis. Innovations in Clinical Neuroscience. 2017; 14: 3–22.

44. Kott A, Daniel D. Understanding Factors Impacting on CGI-S vs. In: CGI-I Discrepancies An Exploratory Analysis Poster presented at the 2015 American society of Clinical Psychopharmacology (ASCP) Annual Mtg. Scottsdale, Arizona.

45. Kott A, Daniel D. S246. The effect of incorrect scale administration on data quality in schizophrenia clinical trials. Schizophrenia Bulletin. 2020; 46: 132–132.

46. Daniel DG, Kott A. Risk-based Monitoring for Aberrant Rating Patterns and Patient Selection Anomalies in Trials. In: Schizophrenia International Research Society (SIRS). Florence Italy. 2014.

47. Echevarria B, Liu C, Negash S, Opler M, Molero P, Capodilupo G. M42. Independent reviewand monitoring improves quality of PANSS datain global clinical trials. Schizophrenia Bulletin. 2020; 46: 150.

48. Daniel DG, Busner J, McNamara C. Ongoing Monitoring and Feedback Decreases Error Rates and Improves Internal Consistency of PANSS Ratings in an International Clinical Trial. In: International Congress on Schizophrenia Research (ICOSR) in Colorado Springs. Colorado; April 2-6 2011.

49. Kott A, Daniel DG. Effects of PANSS audio/video recordings on the presence of identical scorings across visits. Eur Neuropsychopharmacology. 2015; 25: S543–S544.

50. Daniel D, Wang X, Kott A. eCOA Increases Data Quality Compared to Audio/video recording alone in schizophrenia clinical trials. In: International Society for CNS Clinical Trials and Methodology (ISCTM). Washington, DC. 2020.

51. Daniel D, Kott A. Comparison of Audio vs. Audio-Video Recording for Data Quality Monitoring of the Positive and Negative Syndrome Scale (PANSS) 30th European College of Neuropsychopharmacology ECNP Congress, September 2-5, 2017. Paris France

52. Targum SD, Murphy C, Breier A, Brannan SK. Site-independent confirmation of primary site-based PANSS ratings in a schizophrenia trial. J Psychiatr Res. 2021; 144: 241–6.

53. Targum SD, Pendergrass JC, Murphy C. Audio-digital recordings to assess ratings reliability in clinical trials of schizophrenia. Schizophr Res. 2021; 232: 54–60.

54. Kott A, Brannan S, Murphy C, Targum S, Daniel D. Procedures to Optimize Endpoint Data Quality in an Acute Schizophrenia Trial. In: International Society for CNS Clinical Trials and Methodology (ISCTM) Fall, Virtual Conference. September 21-25 2020.

55. Daniel D, David MD, Xingmei W, Kott A. MUDr.: The utility of enhanced data surveillance on the presence of data quality concerns in acute schizophrenia clinical trials. A post hoc analysis. In: International Society for CNS Clinical Trials and Methodology (ISCTM) Autumn Conference [Internet]. Barcelona, Spain; October 5-7 2023.

56. Harvey PD, Miller ML, Moore RC, Depp CA, Parrish EM, Pinkham AE. Capturing Clinical Symptoms with Ecological Momentary Assessment: Convergence of Momentary Reports of Psychotic and Mood Symptoms with Diagnoses and Standard Clinical Assessments. Innov Clin Neurosci. 2021; 18: 24-30.

57. Kott A, Curtin D, Chen Tackett Z, Daniel D. Exploring the Utility of Data Analytics for Identification and Management of Data Quality Concerns in ePRO Data. In: 2019 American Society for Clinical Psychopharmacology (ASCP) Meeting. Scottsdale, Arizona. 2019.

58. Buffenstein I, Kaneakua B, Taylor E. Demographic recruitment bias of adults in United States randomized clinical trials by disease categories between 2008 to 2019: a systematic review and meta-analysis. Sci Rep. 2023; 13: 42.

59. Crittenden-Ward K, Micaletto M, Olt J, Tackett ZC, Machizawa S, Owuor N, et al. Diversity and disparities in research studies and career trajectories in psychiatry. Psychiatry Res. 2022; 308: 114333.

60. Kott A, Wang X, Pintilii E. Andrei Iacob: Using Machine Learning to Identify at Risk Sites in Acute Schizophrenia Clinical Trials. Journal for Clinical Studies. 2022; 14: 24–6.

61. Kott A, Iacob A, Pintilii E, Ariton D, Daniel DG. Assessing temporal performance of machine learning pipelines predicting data quality. In: 2019 American Society for Clinical Psychopharmacology (ASCP). Miami, Florida. 2023.

62. Boesen K, Gøtzsche PC, Ioannidis JPA. EMA and FDA psychiatric drug trial guidelines: assessment of guideline development and trial design recommendations. Epidemiol Psychiatr Sci. 2021; 30: e35.

63. McClellan J, Stock S. American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter for the assessment and treatment of children and adolescents with schizophrenia. Journal of the American Academy of Child & Adolescent Psychiatry. 2013; 52: 976-990.

64. Busner J. Challenges in child and adolescent psychopharmacology clinical trials. Child and Adolescent Psychopharmacology News. 2013; 18: 1–4.

65. Findling RL, Youngstrom EA, Frazier JA, Sikich L, Daniel DG, Busner J. An optimized version of the Positive and Negative Symptoms Scale (PANSS) for pediatric trials. Journal of the American Academy of Child and Adolescent Psychiatry. 2023; 62: 427–34.

66. Busner J, Daniel DG, Findling RL. Identification of PANSS items of particular challenge to raters in adolescent schizophrenia clinical trials. In: Presented as a poster at the Autumn Conference of the International Society for CNS Clinical Trials Methodology. Philadelphia, PA; September 30-October 2 2013.

67. Busner J, Daniel DG, Findling RL. Identification of PANSS items of particular challenge to raters in adolescent schizophrenia clinical trials: Expansion of initial findings. In: 2021 Society for International Research in Schizophrenia (SIRS) (virtual) annual meeting. 2021.

68. Findling RL, Youngstrom EA, Frazier JA, Sikich L, Daniel DG, Busner J. An optimized version of the Positive and Negative Symptoms Scale (PANSS) for pediatric trials. Journal of the American Academy of Child and Adolescent Psychiatry. 2023; 62: 427–34.

69. Youngstrom EA, Langfus JA, Busner J, Daniel DG, Findling RF. Reliability and Accuracy of Scores Using 20 or 10-item short forms of the Positive and Negative Symptoms Scale (PANSS) Replicated in 3 Large Outpatient Trials. In: Accepted as an oral presentation for the American Academy of Child and Adolescent Psychiatry (AACAP) Annual Meeting. New York, NY; 2023.

70. Busner J, Youngstrom EA, Langfus JA, Daniel DG, Findling RL. b) Replicating and extending the reliability, criterion validity, and treatment sensitivity of the PANSS10 and PANSS20 for pediatric trials. In: Presented as a poster presentation at the American Society of Clinical Psychopharmacology (formerly NCDEU) annual meeting. Miami, Florida. 2023.

71. Busner J, Daniel DG, Atkinson SD, Findling RL. a) Development of a child and adolescent eCOA guided interview for the PANSS interview for assessing psychosis in clinical trials. In: Presented as a poster at the American Society of Clinical Psychopharmacology (ASCP), formerly NCDEU, Annual Meeting. Miami, FL. 2020.