Video vs audio-based PANSS Independent Review: A comparison of score discrepancy rates in 2 large acute schizophrenia clinical trials.

Authors: Echevarria, B[1], Sharma, A[1], Negash, S[1], Tatsumi, K[1], De Vito, L[1], Bodelan, M[1] and Opler, M[1]

[1]Clario, a part of Thermo Fisher Scientific

Methodological Issue Being Addressed

Independent review (IRev) of site-administered recordings is often used in schizophrenia clinical trials for quality assurance, continuous training, and increased endpoint data accuracy. Both video and audio recordings of site-administered assessments are used for ascertainment of site scores, but the data comparing the utility of both modalities is limited and largely unexplored.

Introduction

The utility of using both audio and video recording for independent review of the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987) assessments has been previously documented (Targum et al., 2021, Targum et al., 2024); however, it is unclear how use of recording modality impacts score discrepancy rates. We compared two similar studies of acute schizophrenia conducted in the United States (US) and evaluated rates of score discrepancies across recording modalities overall, within PANSS subscales, across items, and within ranges of symptom severity stratified by PANSS total score. Implications of modality selection and impact of visual observation on independent review outcomes are discussed.

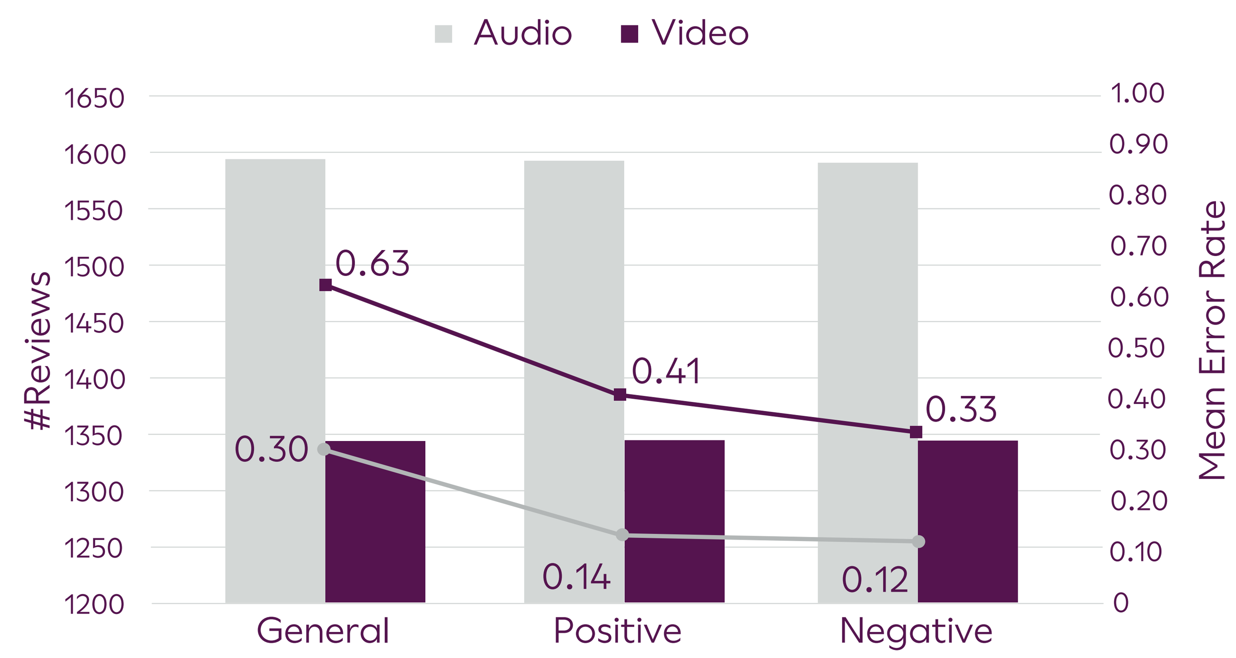
Methods

A total of 2,937 reviewed PANSS assessments from two similar large US multi-site acute schizophrenia clinical trials were compared based on recording modality used for the review (video or audio). Mean error rates resulting from IRev were calculated across recording modalities. Independent sample T-tests were used to assess difference in overall discrepancy rates as well as by PANSS subscale and PANSS total score groups. We also evaluated score discrepancies at the item level.

Results

IRev conducted using video recording (N = 1,593) resulted in higher rates of discrepancies across all evaluated parameters compared to audio (N = 1,344). Overall mean error rates were higher for video than audio ($t(2178.2) = 14.05$, $p < .001$, $d = 0.54$) (Figure 1).

Figure 2. Mean Error Rates by Subscale for Audio & Video


Audio
Video
1650
1600
1550
1500
1450
1400
1350
1300
1250
1200
#Reviews
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
Mean Error Rate
0.63
0.41
0.33
0.30
0.14
0.12
General
Positive
Negative

Similar trends were observed across PANSS total score severity groups, with the most notable difference in the subject group with PANSS TS <60 ($t(143.9) = 6.84$, $p < .001$, $d = 0.73$) (Figure 3).

Figure 3. Mean Error Rates by PANSS Severity Groups

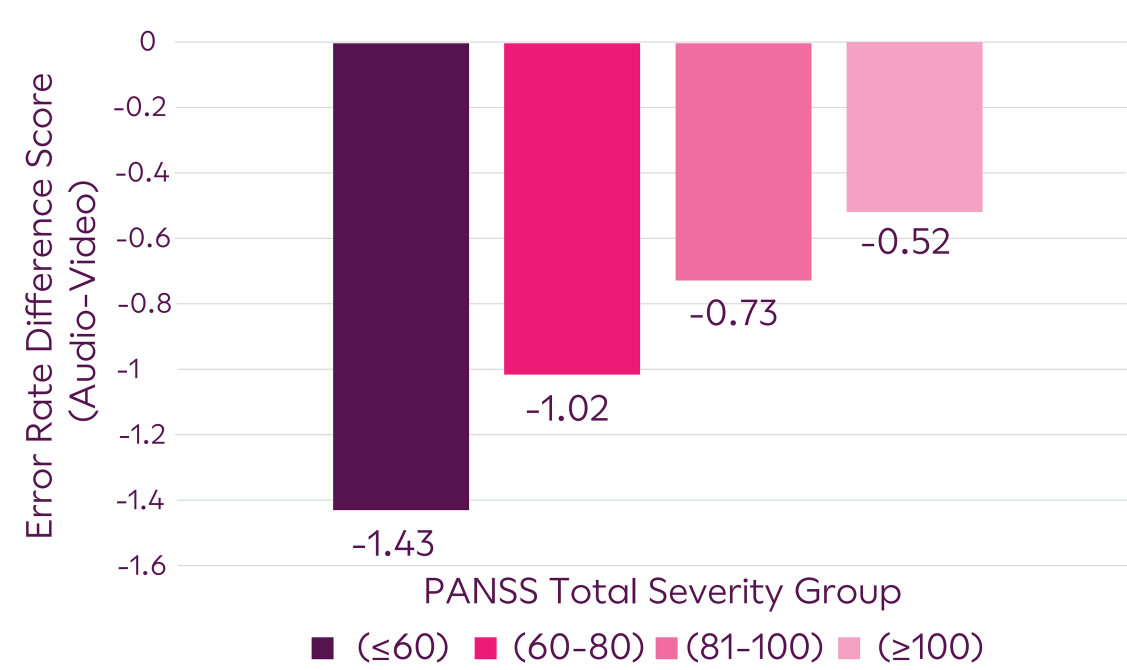
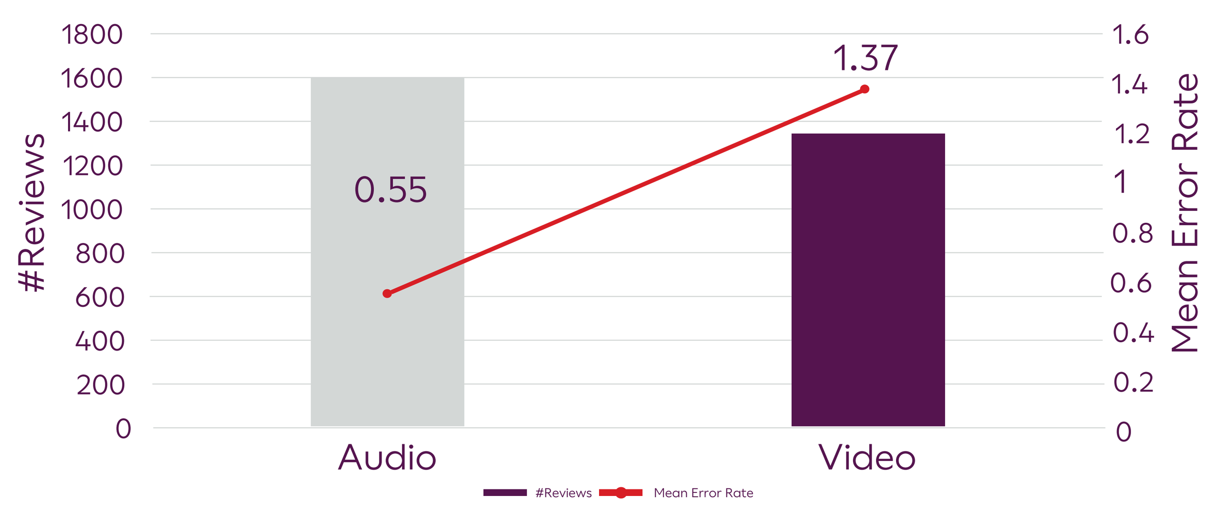

0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
Error Rate Difference Score (Audio-Video)
-1.43
-1.02
-0.73
-0.52
PANSS Total Severity Group
(≤60)
(60-80)
(81-100)
(≥100)

Figure 1. Mean Error Rates by Audio & Video


1800
1600
1400
1200
1000
800
600
400
200
0
#Reviews
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Mean Error Rate
0.55
1.37
Audio
Video
#Reviews
Mean Error Rate

Subscale level analyses also showed significantly higher error rates for video than audio: positive: $t(1989.1) = 12.09$, $p < .001$, $d = 0.47$), negative: $t(2171.9) = 9.55$, $p < .001$, $d = 0.37$), and general: $t(2318.7) = 9.69$, $p < .001$, $d = 0.37$ (Figure 2).

Individual items with highest difference in discrepancy rates include P1, P6, G2, and N1 (Table 1).

Table 1. Mean Error Rates by PANSS Items

Item Name	Audio	Video
P1. Delusions	1.95	8.56
P6. Suspiciousness/persecution	0.88	7.22
G2. Anxiety	2.07	7.14
N1. Blunted affect	0.25	4.69
G12. Lack of judgment and insight	3.70	8.04
N2. Emotional withdrawal	1.76	5.95
P3. Hallucinatory behavior	1.82	5.8
P4. Excitement	1.07	4.91
N4. Passive social withdrawal	7.28	10.86
G16. Active social avoidance	6.15	9.6
G4. Tension	1.19	4.46
P5. Grandiosity	2.13	5.36
G5. Mannerisms and posturing	0.63	3.42
G6. Depression	3.70	6.25
N6. Lack of spontaneity	1.51	4.02
P2. Conceptual disorganization	1.82	4.24
P7. Hostility	2.51	4.91
N5. Difficulty in abstract thinking	2.26	4.61
G7. Motor retardation	0.69	2.68
G9. Unusual thought content	1.82	3.79
N3. Poor rapport	0.13	2.08
G10. Disorientation	1.63	3.50
G8. Uncooperativeness	1.00	2.6
G3. Guilt feelings	0.88	2.16
G15. Preoccupation	0.25	1.34
G1. Somatic concern	3.08	3.87
N7. Stereotyped thinking	0.38	1.12
G14. Poor impulse control	1.44	2.16
G13. Disturbance of volition	0.25	0.82
G11. Poor attention	1.07	1.04

Conclusions

Our findings indicate that conducting IRev using video recordings instead of only audio results in higher rate of score discrepancies across subscales and symptom severity levels. This was found for items that require visual observation for rating (e.g. N1) as well as for items that are less reliant on visual inspection (e.g. P1, P6). The fact that video recordings yield higher score discrepancy rates across all parameters evaluated may reflect increased confidence in reviewer's own scores and rationales derived from visual observation of the subject. The findings suggest increased benefit of using video IRev compared to audio. Since use of video recordings has both cost as well as privacy implications, our results can be leveraged for cost-benefit analyses when deciding which IRev modality to use.

Future studies should aim to replicate this work by expanding the number of studies and into other schizophrenia subgroups. Also, a limitation and potential focus of future work is evaluating the directionality of discrepancies (e.g. whether site or review score is higher) and the impact of IRev discrepancies on study outcome.

References

1. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261-76. doi: 10.1093/schbul/13.2.261. PMID: 3616518.
2. Targum SD, Ge T, Asgharnejad M, Reksoprodjo P, Singh JB, Murthy V. Use of video-recordings of site-based interviews for quality assurance in a study of subjects with schizophrenia and persistent negative symptoms. Schizophr Res. 2024 Oct;272:61-68. doi: 10.1016/j.schres.2024.08.014. Epub 2024 Aug 26. PMID: 39190983.
3. Targum SD, Pendergrass JC, Murphy C. Audio-digital recordings to assess ratings reliability in clinical trials of schizophrenia. Schizophr Res. 2021 Jun;232:54-60. doi: 10.1016/j.schres.2021.05.004. Epub 2021 May 17. PMID: 34015556.

Disclosures
The authors report no conflicts of interest for this work

CLARIO.