International
CLINICAL TRIALS

Decentralising clinician-administered assessments: evidence, adaptation and oversight

Psychiatric and neurological rating scales were built for face-to-face administration, but the trials evaluating them increasingly are not. This article examines the evidence, the practical adaptations and the implementation models that make remote clinician-reported outcome measures assessments work without compromising data quality

Bill Byrom, Marcela Roy and Rachel Berman at Signant Health

The era of the decentralised clinical trial (DCT) conducted entirely outside traditional research sites proved shorter-lived than many anticipated. While the pandemic-era urgency accelerating DCT adoption has passed, its legacy is a pragmatic hybrid model: site-run trials that strategically incorporate decentralised elements to reduce patient burden and broaden access. Today, remote assessments are increasingly standard features of conventionally run studies, rather than hallmarks of a distinct trial category.

Within this evolving landscape, clinician-administered outcome measures (ClinROs) warrant particular attention. Instruments such as the Montgomery-Åsberg Depression Rating Scale (MADRS) and Hamilton Depression Rating Scale (HAM-D) have long been the gold standard for efficacy assessment in psychiatric and neurological trials, yet they were designed for face-to-face in-person administration.

As sponsors and sites explore conducting these assessments remotely via video or telephone, critical questions arise around measurement equivalence, rater standardisation and regulatory acceptability. This article examines the evidence base and practical considerations for remote ClinRO administration in trials incorporating decentralised assessments.

Three distinct implementation models will be addressed: assessments conducted by site-based raters using familiar remote-visit workflows; centralised raters administering scales independently of the enrolling site; and central scorers rating from recorded or transcribed assessment interviews conducted by site raters. Each model carries different implications for data quality, operational complexity and trial integrity, and the appropriate choice depends on the scale, indication and broader study design.

Evidence and rationale

The evidence base for remote ClinRO administration predates the pandemic-era shift towards trials with decentralised elements, with foundational work focused on structured, interview-based psychiatric scales most used as primary endpoints.

For the HAM-D, Kobak found that mean scores were comparable across face-to-face and video conference modalities, with an intraclass correlation of 0.88, and that inter-rater reliability between video and in-person administration was not significantly different.[1] The MADRS showed a similar performance, with intraclass correlation coefficients (ICCs) of .94 and .93 for video conference and telephone administration respectively, comparable to the test-retest reliability seen with two face-to-face administrations.[2]

Neurological scales follow a similar pattern with scale-specific caveats. Tai et al demonstrated excellent test-retest reliability for the mFARS (ICC = 0.97) and SARA (ICC = 0.98) by video in Friedreich's ataxia, though reliability was lower for the upper limb coordination subscale, attributed to the difficulty of assessing fine finger movements via video.[3] Chapman et al found no significant difference in total Montreal Cognitive Assessment (MoCA) scores between face-to-face and video conference administration in stroke survivors, though wide limits of agreement at the individual level are a reminder that equivalent group means can mask meaningful variability in individual patient scores.[4]

Regulatory frameworks have moved broadly in the direction of endorsing remote assessment. The US Food and Drug Administration's final guidance, '*Conducting Clinical Trials with Decentralized Elements*', explicitly endorses telehealth visits as a mechanism for conducting trial-related activities remotely.[5] Although the guidance does not address ClinRO administration specifically, the signal is clear: remote assessment is acceptable in principle, provided sponsors demonstrate measurement equivalence and document rater qualifications.

Measure suitability and adaptation

Remote administration works best for scales with clearly defined constructs, verbally elicited responses and minimal reliance on equipment or tightly controlled settings. Interview-based ClinROs such as the MADRS and HAM-D, and measures using simple gross motor observations, tend to adapt well. Performance-based scales are harder to migrate and may require item omission or validated revised scoring.

Several flagship scales illustrate feasible adaptations. The Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part Three has been implemented via video with structured guides, enabling rating of many motor items, while omitting elements like the pull test and rigidity and interpreting scores as 'partial' motor exams.[6] A modified remote National Institutes of Health Stroke Scale (rNIHSS) has been developed by dropping low-reliability items to improve remote agreement.[7] The Telephone MoCA removes visuospatial and drawing tasks, and can discriminate mild cognitive impairment from cognitively normal older adults, though should not be considered interchangeable with the full MoCA.[8]

Any remote adaptation is effectively a scale modification and requires evidence that scores remain meaningful in the new modality, typically through qualitative testing and quantitative comparisons between remote and in-person administration. If equivalence cannot be shown, the remote version should be treated as a related but distinct instrument.

Content that does not decentralise well includes tasks requiring physical contact between the rater and subject, specialised hardware, or subtle motor or facial signs that can be lost with poor video quality.[9] The presence

of caregivers or others near the interviewee can introduce coaching and expectancy bias and requires procedural controls to minimise these potentially confounding factors.

Implementation models: site raters; central raters; central scorers

In trials with decentralised elements, clinician-administered assessments must work reliably when patients are at home, at community sites or in mixed visit schedules. Site raters, central raters and central scorers can each support remote or partially remote models, and sponsors often combine them to balance feasibility with data quality.

Site raters in decentralised assessments are local investigators or research staff who conduct televisits as part of the trial alongside traditional site visits. This model has been deployed in Alzheimer's disease (AD) and schizophrenia programmes where investigators integrate ratings with longitudinal clinical management. Central quality monitoring acts as an overlay, using near real-time analytics to identify unexpected patterns across dispersed sites, for example, shifts in Positive and Negative Syndrome Scale (PANSS) distributions in a largely remote schizophrenia study or unusual learning effects on AD cognitive batteries. Outlier sites can be targeted for retraining or review without disrupting the decentralised visit model.[10]

Central raters are remote expert clinicians who perform audio or video-based assessments, naturally compatible with decentralised designs because rating is decoupled from the physical trial location. In psychedelic major depressive disorder (MDD) trials, for example, teleconference-based central rating structurally separates treatment administration from outcome assessment, helping minimise functional unblinding in the face of prominent subjective drug effects. Trials of psilocybin for MDD have successfully used remote, blinded raters for MADRS as the primary endpoint, with raters kept unaware of treatment assignment, visit number and protocol details.[11,12]

Central scorers are well suited to decentralised neurology trials that still require hands-on exams at local sites. In Parkinson disease, for example, participants may attend nearby clinics for MDS-UPDRS motor exams, while video is captured and sent to centrally trained scorers who apply standardised scoring criteria.[13] This allows sponsors to leverage local access where appropriate, while central scorers reduce between-site scoring noise that could obscure modest treatment effects.

Quality control and surveillance

Quality control must account for the increased geographic dispersion, varied visit types and technology reliance inherent in DCTs. For trials using decentralised site raters, statistical surveillance of rating distributions, within-patient change and scale use patterns can highlight raters whose assessments deviate from peers, triggering focused review of interview recordings, feedback and refresher training. This approach respects local clinical relationships while addressing risks introduced by remote visits, staff turnover and uneven experience with tele-assessment.

For trials implementing central raters, surveillance focuses on maintaining calibration and blinding.[11] Regular calibration sessions using shared reference videos, double rating of a subset of remote interviews and periodic inter-rater reliability checks help ensure consistent scoring over time and across global rater pools. Monitoring raters' guesses about treatment assignment can further confirm that central rating is

mitigating functional unblinding rather than amplifying local expectations.

For trials using central scoring, standardised protocols for capturing home or community clinic videos and review by multiple central scorers support reliable MDS-UPDRS scoring.[14] Inter-rater reliability metrics and periodic consensus sessions help align scorers as remote video quality or patient environment varies across decentralised sites.[15] When available, AI or digital motion analyses can be layered on to flag discordant patterns, which experts then adjudicate, combining scalable monitoring with clinical oversight.

Conclusion

Thoughtfully implemented decentralised ratings can streamline visits, expand access, increase rater reliability and support more patient-centric designs without sacrificing scientific rigour. By grounding model choice and scale adaptation in modality-specific validation, risk-based surveillance and scale-by-scale evidence, sponsors can make defensible, evidence-based decisions that enhance efficiency while preserving trial integrity.

References:

1. *Visit: pubmed.ncbi.nlm.nih.gov/15273034/*
2. *Visit: pubmed.ncbi.nlm.nih.gov/17941100/*
3. *Tai G et al (2021), 'Determining the validity of conducting rating scales in Friedreich ataxia through video', Movement Disorders Clinical Practice, 8(5):688-693*
4. *Chapman JE et al (2021), 'Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke', Journal of Telemedicine and Telecare, 27(8):484-492*
5. *Visit: fda.gov/regulatory-information/search-fda-guidance-documents/conducting-clinical-trials-decentralized-elements*
6. *Visit: onlinelibrary.wiley.com/doi/10.1155/2016/4802570*
7. *Visit: sciencedirect.com/science/article/abs/pii/S0303846725003488?via%3Dihub*
8. *Wong A et al (2015), 'Montreal Cognitive Assessment 5-minute protocol in stroke and transient ischemic attack: Evaluating its psychometric properties and relationship with functional outcomes', Stroke, 46(4), 1059-1064*
9. *Visit: pearsonclinical.co.uk/digital-solutions/telepractice/telepractice-and-the-wais-iv.html?srsltid=AfmBOorJi8ShbJpbANe-BL2TRhtdpCduD8b4b9tngsBeIQcAdolmOqW_*
10. *Visit: pubmed.ncbi.nlm.nih.gov/34700212/*
11. *Raison C L et al (2023), 'Single-dose psilocybin treatment for major depressive disorder: A randomized clinical trial', JAMA, 330(9), 843–853*
12. *Visit: nejm.org/doi/full/10.1056/NEJMoa2206443*
13. *Visit: journals.sagepub.com/doi/10.3233/JPD-202163*
14. *Visit: nature.com/articles/s41531-020-00135-w*
15. *Visit: movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mdc3.14203*

Bill Byrom PhD is vice president of Product Intelligence and Positioning at **Signant Health**. Bill is a recognised leader in eClinical product strategy and electronic clinical outcome assessments with over 30 years of pharmaceutical industry experience. He combines deep clinical development expertise with technological innovation to advance digital health solutions. Bill has authored over 80 publications and two industry textbooks on electronic patient-reported outcomes. He consults with pharmaceutical companies and regulatory agencies on digital health initiatives and serves on various industry working groups and consortia.

Marcela Roy MA is executive director in Clinical Science and Medicine at **Signant Health**. She has been with Signant for over 15 years and has over 20 years of clinical and research experience. Her focus is on mood disorders and endpoint reliability, with a strong emphasis on quality monitoring in DCTs. She provides strategic direction within the organisation and leads cross-functional teams to advance the design, operational excellence and scientific impact of patient-centric, decentralised research, working closely with sponsors and partners to support innovative trial execution.

Rachel Berman PhD brings two decades of neuropsychology experience to **Signant Health**, where she advises sponsors on clinical trial design, rater training, and endpoint monitoring in psychiatric and neurological disorders, including schizophrenia, mood disorders, rare diseases, and neuropsychiatric and cognitive symptoms of dementia. She also lends her expertise to scientific and strategic design considerations for psychedelic trial implementation. Alongside her clinical science and research leadership, Rachel also collaborates on business development initiatives, ensuring scientifically grounded solutions align with sponsor needs and study objectives.