Title

Remote Assessment of Negative Symptoms of Schizophrenia.

Journal

Schizophrenia Bulletin Open, 4(1)

Authors

Daniel, David

Cohen, Alex

Velligan, Dawn

et al.

Publication Date

2023

DOI

10.1093/schizbullopen/sgad001

Peer reviewed

Schizophrenia Bulletin Open
https://doi.org/10.1093/schizbullopen/sgad001

MAJOR REVIEW (ORIG)

Remote Assessment of Negative Symptoms of Schizophrenia

David G. Daniel[*,1], Alex S. Cohen[2], Dawn Velligan[3], Phillip D. Harvey[4,5], Larry Alphs[6], Michael Davidson[7], William Potter[8], Alan Kott[9], Nina Schooler[10], Christopher R. Brodie[11], Raeanne C. Moore[12], Pierre Lindenmeyer[13], and Stephen R. Marder[14]

[1]Signant Health, Mclean, VA, USA; [2]Louisiana State University, Baton Rouge, LA, USA; [3]University of Texas Health Science Center at San Antonio, San Antonio, TX, USA; [4]University of Miami, Miami, FL, USA; [5]Research Service, Bruce W. Carter VA Medical Center, Miami, FL, USA; [6]Denovo BioPharma, San Diego, CA, USA; [7]Minerva Neurosciences, Inc, Burlington, MA, USA; [8]Independent Expert, Philadelphia, PA, USA; [9]Signant Health, Prague, Czech Republic; [10]SUNY Downstate Medical Center, Brooklyn, NY, USA; [11]Otsuka Pharmaceutical Development and Commercialization, Inc, Princeton, NJ, USA; [12]University of California San Diego, San Diego, CA, USA; [13]Nathan S Kline Institute for Psychiatric Research, Orangeburg, NY, USA; [14]Semel Institute for Neuroscience at UCLA and the VA Desert Pacific Mental Illness Research, Education and Clinical Center, Los Angeles, CA, USA

[*]To whom correspondence should be addressed; Signant Health, PO Box 7137, McLean, Va 22106, USA; tel: 703-638-2500, e-mail: dgdanielmd@gmail.com

In contrast to the validated scales for face-to-face assessment of negative symptoms, no widely accepted tools currently exist for remote monitoring of negative symptoms. Remote assessment of negative symptoms can be broadly divided into 3 categories: (1) remote administration of an existing negative-symptom scale by a clinician, in real time, using videoconference technology to communicate with the patient; (2) direct inference of negative symptoms through detection and analysis of the patient's voice, appearance, or activity by way of the patient's smartphone or other device; and (3) ecological momentary assessment, in which the patient self-reports their condition upon receipt of periodic prompts from a smartphone or other device during their daily routine. These modalities vary in cost, technological complexity, and applicability to the different negative-symptom domains. Each modality has unique strengths, weaknesses, and issues with validation. As a result, an optimal solution may be more likely to employ several techniques than to use a single tool. For remote assessment of negative symptoms to be adopted as primary or secondary endpoints in regulated clinical trials, appropriate psychometric standards will need to be met. Standards for substituting 1 set of measures for another, as well as what constitutes a "gold" reference standard, will need to be precisely defined and a process for defining them developed. Despite over 4 decades of progress toward this goal, significant work remains to be done before clinical trials addressing negative symptoms can utilize remotely assessed secondary or primary outcome measures.

Key words: schizophrenia/negative symptoms/remote assessment

Introduction

Negative symptoms represent an important target for intervention research in schizophrenia. They are strongly related to functional outcomes, yet the available pharmacological treatments for negative symptoms are ineffective compared to the improvement in positive symptoms with antipsychotics.[1] Researchers continue to reexamine existing drugs and design novel ones that target negative symptoms. At the same time, a handful of groups are working to develop assessment tools to measure more effectively the changes in negative symptoms associated with various interventions. This effort is still in its early phases; unlike the positive symptoms of psychosis, which can be defined and measured in a relatively straightforward manner, negative symptoms defy easy characterization and quantitation.

Part of the difficulty in developing measurements for negative symptoms stems from the lack of a pathophysiological mechanism. In its absence, psychometric analyses of schizophrenia-related behaviors have not clearly resolved the question of which features of the disease should be included under the umbrella of negative symptoms. For example, social withdrawal may be secondary to delusions and hallucinations (ie, positive symptoms), or it may be independent of such psychosis

(ie, negative symptoms). Thus, the interpretation of asociality as a positive or negative symptom depends on context and therefore is a more-subjective assessment than is ideal. Furthermore, therapeutic interventions might resolve some negative symptoms but leave others untouched, which could cause an assessment tool to appear insensitive when in fact it was merely incomplete.

Despite these hurdles, several instruments now exist to quantify negative symptoms during a face-to-face interview with the patient. These rating systems include the 16-question Negative Symptom Assessment (NSA-16), the Brief Negative Symptom Ratings Scale (BNSS), and the Clinical Assessment Interview for Negative Symptoms (CAINS).[2–5]

In contrast to the validated scales for face-to-face assessment of negative symptoms, no widely accepted tools currently exist for remote monitoring of negative symptoms. Yet the advantages of such technologies could be substantial. In addition to the potential benefits of accuracy and convenience, digital/virtual assessments might help address some of the current COVID-related constraints on patient care and continuing research. It is even conceivable that certain clinical features might be better quantified with digital tools than with interview-based metrics.

Remote assessment of negative symptoms can be broadly divided into 3 categories: (1) remote administration of an existing negative-symptom scale (eg, NSA-16 and BNSS) by a clinician, in real time, using audio/video (videoconference) technology to communicate with the patient; (2) direct inference of negative symptoms through detection and analysis of the patient's voice, appearance, or activity by way of the patient's smartphone or other device; and (3) ecological momentary assessment (EMA), in which the patient self-reports their condition upon receipt of periodic prompts from a smartphone or other device computer interface during their daily routine. Each modality has unique strengths, weaknesses, and issues with validation. As a result, an optimal solution may be more likely to employ several techniques than to use a single tool. The specific combination will depend on which domain or domains are being targeted by a given intervention.

These modalities vary in cost, technological complexity, and applicability to the different negative-symptom domains. For example, EMA may be a better tool for reporting the subjective perception of reduced emotional experience, whereas computerized facial and vocal analysis may be better suited to objective measurement of diminished affective expression. In the case of complex negative-symptom domains, such as diminished goal-directed behavior and interest, a combination of techniques—perhaps movement tracking (actigraphy) and EMA—may provide a clearer clinical picture than either metric alone. Some assessment techniques are passive or automated, whereas others require the patient's cooperation in the form of recording behaviors, answering surveys, transmitting responses, and charging and maintaining the devices. Finally, each technique or combination of techniques requires validation. As an added layer of complexity to the validation process, it is possible, even likely, that some of the remote measures will assess features of schizophrenia that are not captured by the currently available face-to-face scales. Although such metrics may lack a gold standard for validation, they are no less important. These novel assessments could yield new insights to further the mechanistic and clinical understanding of the disease.

Remote Administration of Negative-Symptom Rating Scales Utilizing Audio-Video Technology

Both newer and older scales were developed and validated with in-person interviews. As the field of clinical trials moves toward remote assessment, it is important to understand how these instruments perform in these new conditions. One result of restrictions related to the COVID-19 epidemic is that many clinical trials found it necessary to move to remote assessment. In some cases, both investigators and subjects found that there are advantages associated with video assessment in terms of convenience and cost. By necessity, study teams mastered the technology associated with video rating and assumed—with limited evidence—that clinical ratings were valid.

However, it is reasonable to speculate that negative symptoms may represent a domain of psychopathology that does not easily transfer to video. For example, positive symptoms such as frightening delusions or auditory hallucinations may be relatively straightforward for patients to describe, whereas negative symptoms are likely to be rated based on the interactions of the subject with the rater. Interactions between a rater and a patient may lead to changes in a subject's responses which, in turn, will result in rated changes in responsiveness that are unrelated to treatment effects. In addition, especially in the inpatient setting, the person carrying out an in-person interview can make observations that occur before and after the interview that are not available to a remote interviewer. Remote evaluations can be influenced by "remote variables" including, eg, noise, interpersonal conflicts, pressure and other distressing or distracting facets of the home environment. Another challenge for remote assessment is that negative-symptom patients who are unable or unwilling to master the technology of remote assessment may not be represented in these trials.

There are also potential advantages for remote assessment of negative symptoms. If a patient is observed during a remotely conducted video interview at home, there may be information from the setting that will be valuable in assessing the person's day-to-day activities. Whereas in-person administration of the negative-symptom factor

of the Positive and Negative Syndrome Scale (PANSS), NSA-16, BNSS, and CAINS in the clinical setting have for the most part demonstrated satisfactory psychometric performance, reports of remote assessment in psychotic disorders—while encouraging—are very sparse.[2,4–8] Full psychometric evaluation of these scales administered remotely to schizophrenia patients with prominent symptoms is for the most part still pending. Myriad factors inherent to evaluating a subject by videoconference in the home environment may impact the subject's forthrightness, behavior, and appearance and cause remote and in-person measurement of negative symptoms to differ. The interview structure, teleconference devices, camera view, transmission bandwidth, setting (preferably quiet and private), and informant, if available, should be kept constant across visits.

A critical step in validating remote administration of clinician-administered negative-symptom scales is comparison to in-person administration in the same subjects, ideally conducted by a pair or pool of well-calibrated raters who are blinded to each other's ratings. The alternating in-person vs remote assessments should be conducted relatively close in time for subjects whose symptomology may be changing due to a pharmacological or other intervention. In stable patients, longer intervals between remote and clinic assessment would be less likely to confound comparison. Intraclass correlation coefficients may be calculated between the in-person and remotely administered scale assessments. Reliability of remote vs in-person measurement of change from baseline in schizophrenia symptoms should also be assessed.

Data analytic assessment of the expected relationship among items within and across scales is a parsimonious methodology for ongoing surveillance of data quality of remote administration of clinician-administered scales. Periodic comparison to in-person administration should also be conducted for quality assurance purposes. Random and for cause samples of collected interview recordings can be blindly reviewed by independent evaluators to identify possible errors in scale administration and scoring.

EMAs, which frequently question the subject, and digital measures that record behaviors in real time are emerging exploratory methodologies that may provide further insight into the validity of standard clinical assessments administered remotely. The latter are subject to recall inaccuracy and bias because they typically require a subject to report their mental status and behaviors retrospectively for a 1-week look-back period.

In summary, remote clinician administration of negative-symptom scales is technologically feasible but as yet unvalidated. It has the potential advantages of assessing patients in their home environment and avoiding logistic challenges of office visits. However, the remote technology may be challenging or uncomfortable to some patients, be difficult to standardize, and affect interpersonal behaviors in ways that impact rating differently from in-person assessments. In patient populations other than schizophrenia, comparability of remote and in-person assessment of cognition, behavior, and mood have been demonstrated on some measures, but validation and norms are lacking.[9,10]

Remote Assessment of Negative Symptoms Utilizing Computerized Audio-Video Analysis

Audio-video media, whether procured from a traditional "face-to-face" clinical interview, from a virtual interview, or from a variety of video selfie, social media, and other formats, can be analyzed using a variety of facial, natural language, vocal, and other computerized analytic approaches. This can provide important information about blunted affect, alogia, social motivation, hedonic experience, and interest.

Audio-video analysis has the benefit of objectivity and sensitivity. Computerized facial analysis, eg, can objectively assess facial blunting/expression and has the advantage of being able to detect more subtle changes in facial movement than those that can be measured by a rater conducting a clinical interview.[11,12] In addition, this method avoids variance associated with rater bias and deficits in intra- and inter-rater reliability. Technologies can identify movement of facial features and integrate these movements into corresponding affects using algorithms or can examine discrete facial movements without specifying the type of affect they may represent, focusing instead on quantity and intensity of movements. Convenience and cost are other obvious advantages of these technological strategies, as data collection, processing, analysis, and interpretation can be automated to various degrees. Also, patients might be less anxious and more open to disclose feelings and concerns in a familiar environment. This methodology avoids interactions between a rater and a patient that may lead to changes in a subject's response, which in turn, could result in rated changes in responsiveness that are not due to treatment effects. Finally, there are some interactions and behaviors related to daily activities that may be specific to the home environment.

On the downside, audio-video analysis often requires technological competence by patients, and some patients could be intimidated by the remote assessment, and technological problems such as poor connections might get in the way. Moreover, they do not permit "hypothesis testing" as needed for many semistructured interviews, as live interviewers are able to shape the interview in ways that help them gather important confirmatory or disconfirmity evidence.

Ecological Momentary Assessment

Patients engaged in EMA assessment reply at predefined intervals to specific questions regarding their momentary experiences of emotions, interest, motivations,

and engagement in various activities. In most face-to-face assessments, activities and emotions are rated based upon the participants' recall over the previous days or week(s).[13,14] These momentary assessments have been found to be less impacted by memory impairments in comparison to retrospective recall of emotion, which is required for traditional rating interviews.[15] Studies have shown that individuals with schizophrenia overrate the intensity of experienced emotions when recalling these events during in-office ratings.[16] Ratings of the momentary experiences of emotions captured by EMA have been found to distinguish between control and patient groups.[14] Adherence in wide-ranging samples tends to run over 75% with compensation.[17] Missing data may be addressed by application of maximum likelihood methods for cases who meet the overall adherence criterion.

Evaluation of avolition illustrates many of the potential challenges and opportunities of EMA assessment of negative symptoms. Avolition is typically defined as reduced subjective and objective motivation to engage in pleasurable and essential activities and is a central component of negative symptoms and is a critical target for pharmacological and nonpharmacological intervention.[18,19] However, because of its complexity and the plethora of external factors affecting avolition, it is challenging for an individual to report this experience based on summaries of recalled events and emotions in face-to-face interviews. Other things equal, being home (vs away), alone (vs with someone, particularly someone of your own choosing), and engaging in relatively more unproductive activities (pacing, smoking, watching TV, resting, sitting alone, doing nothing) would be evidence for greater severity of avolition, particularly if accompanied by the subjective lack of motivation to do anything different. Using EMA to ask the simple question: "Where are you?" is a valuable strategy that has been shown to correlate with clinical ratings of negative symptoms.[13,20,21] The proportion of surveys answered with "home" as an outcome provides a simple but powerful index of location. Similarly, participants can momentarily quantify the amount of time at home in a predetermined time period. Activities have been surveyed on a momentary basis and higher levels of engagement in particularly unproductive activities indexes the behavioral features of avolition.[22,23] Subjective report can be complemented with analysis of Global Positioning Systems (GPS) coordinates, which are collected with most smartphone and social media applications. Geolocation can be centered on the home and passively measure the number of times the participant leaves the home, how far they go, and how long they stay away.[14,24,25] Studies cited above found excellent convergence between self-reported and GPS measured locations, with both streams correlating with clinician ratings of negative symptoms. Interpretation of GPS-based analyses may be complicated by cultural, pandemic, and other environmental factors.

Actigraphy, involving measuring movement using wearables, smartphone sensors, home cameras, and other devices, can also be used to measure productive activities, including movement and sleep, as well as sleep at unproductive times. Raugh et al reported on the feasibility and validity of multichannel assessment of actigraphy, GPS, and EMA, finding that answering up to 10 EMA prompts per day was feasible and that both active and passive remote assessments were performed commonly enough to generate valid data.[24] The combination of EMA, actigraphy, and GPS data can be used for convergence regarding the various correlates of avolition and other negative symptoms. For example, Strassnig et al found that people with schizophrenia were more likely to report that they were sitting and less likely to be moving than healthy people in a case control study.[22] Further, participants with schizophrenia were more likely to report only a single activity in the past hour. GPS and actigraphy data could be combined with activity EMA probes to capture behavioral elements of avolition reliably and validly. EMA surveys and GPS measurements can separate high levels of steps indexed by actigraphy that are associated with exercise vs. agitation by indexing location and capturing reports of current activities.

Independent of combining with actigraphy, in terms of subjective avolition, EMA is very well suited to capture of momentary experiences in domains of moods, satisfaction, and intentions. Thus, the experience of subjective avolition would be marked by reduced motivation to engage in the activities described above and reduced levels of dissatisfaction with social isolation, unemployment and associated poverty, and simple boredom. For example, Jones et al found that participants with schizophrenia who reported that they were never sad were more commonly reporting that they were home than participants who reported occasional sadness.[26] They also reported that they had greater competence in functional domains than those who were occasionally sad, although there was no difference in objective performance. This is an example of how EMA data can be used to capture momentary states consistent with subjective avolition and their convergence with objective indices of behaviors consistent with avolition.

Similar to avolition, assessment of diminished hedonic drive, per se, in persons with negative symptoms is impacted by numerous factors that affect the validity, reliability, and stability of patient reports across clinical states, whether these assessments are made in person or remotely.

Implicit in the assessment of a patient's current hedonic state is that it represents the internal state of mind of that individual. As such, it is subjective and highly dependent on an internal mindset. From the patient's perspective it is dependent on a capacity for self-reflection which is usually majorly impacted by negative symptoms. More specifically, self-assessment of one's hedonic drive

is dependent on one's frame of reference and particular world view at the time of assessment. Accurate reporting further depends on motivation, energy level, and mental capacity. All of these are impacted by negative symptoms and could be expected to change as the overall severity of negative symptoms is modified.

The potential of self-report instruments to capture various aspects of reward experience has been demonstrated recently.[14,15] Similar to avolition, self-report measures can disentangle temporal components of reward experience, eg, involving evaluation "in the moment" (eg, "how much are you enjoying this activity") as well as prospectively (eg, "how much will you enjoy this activity") and retrospectively (eg, "how much did you enjoy this activity").[27,28] These components have been tied to biologically distinct reward systems, as well as functioning and symptom states.[29–31] As with other domains of negative symptoms remote assessment of anhedonia can reduce retrospective bias, and in doing so, improve specificity to contexts of interest.[32] This is important regarding, among other things, social vs nonsocial contexts (eg, "how much are enjoying this interaction"); a distinction thought critical to schizophrenia-spectrum pathology for over 50 years.[33,34] Relatedly, remote assessment can potentially help tease apart "secondary" (eg, co-occurring social anxiety, depression, and social isolation) from idiopathic causes of anhedonia.[35] Finally, EMA assessment of anhedonia can be personalized such that specific activities of interest can be evaluated based on prior responses (eg, "how much did you enjoy your date with Pat?"). This approach can help evaluate personalized goals and help contextualize anhedonia report. As with avolition, the latter is important given the reality that opportunities to experience positive emotion, and even how it is conceptualized and communicated, varies across individuals, and systematically as a function of culture, and environment.[36]

Like avolition, diminished hedonic drive can also be inferred from behavior, which can be recorded through a variety of media, including language, physical motion, and reaction time.[11,12] For example, positive emotion and anhedonia have been associated with various semantic aspects of language (eg, "My date with Pat was OK, but I didn't really feel a connection") and can be objectified using a variety of automated solutions.[37,38] As with anhedonia, evaluating physical activity using actigraphy sensors, geolocation, or video analysis of body movement can reveal information about anhedonia, and convergence with clinical ratings have been reported in a number of studies.[14,24] Behavioral tasks tapping reward related RDoCs could also potentially be adapted to remote assessment. Laboratory tasks have been developed to capture reward anticipation, probabilistic and reinforcement learning, and reward valuation, and it is reasonable to think these tasks could be meaningfully adapted for remote applications.[39,40] While proof of concept for these technologies has been, or is being evaluated, a comprehensive evaluation of their psychometrics has yet to be established.

Anergia, involving reduced activity and movement, is certainly addressed by all of the technology-based strategies described above. Reduced purposeful and spontaneous movement can be captured passively with actigraphy and the topography of concurrent activities can be indexed with EMA data.[22,23]

There are, however, multiple limitations to implementation of these technologies at the present time. Responses to EMA surveys have been found to vary considerably based upon the activity occurring at the time the sample is taken. This may impact the ability to detect change. For some domains (eg, affective expression), asking participants to audio/video record responses to highly structured interview questions may unnaturally standardize data and make the digital data collected more likely to correspond to data gathered during a clinical interview. Whether this makes digital data collection ratings more or less valid is an open question. Relationships among negative symptoms assessed using technology may be diffuse. For example, video collected facial expressions have also been found to be related to anhedonia, avolition, asociality, motor retardation, and sexual interest. How this would impact a chosen study outcome is not clear. The number and length of samples needed to reliably detect subjective and objective negative symptoms and their changes over time is not clear. The best measures derived from digital data for each domain have not yet been identified (ie, SD vs mean vs various combinations, distances vs numbers and lengths of times away, proportion of surveys answered while experiencing an emotional state vs. subjectively rated intensity of the emotion, etc.) and these may vary across sampling strategies. Algorithms behind some programs used to quantify data may be proprietary and constitute a mysterious black box, making identifying exactly what is produced difficult to interpret and even more challenging to justify to regulatory agencies. Image and voice resolution can impact some digital ratings, suggesting recorded vs streamed data may be the best way to approach standardization for some measures. It is unclear how to combine different measures into a meaningful approximation of negative symptoms. For example, a combination of strategies may have the ability to distinguish activities that could be exercise from those representing agitation. Taking 6000 steps in an hour while the GPS says that you are at home might be a sign of an agitated episode, while the same number of steps when the GPS suggests that you are at a park or a high-school track suggests exercise. Specific and tested algorithms to combine data sources to improve reliability and validity are needed. The generalizability of data collected using norms based on US samples is unclear and larger representative samples across cultures would likely be needed before this technology could be utilized in multisite clinical pharmaceutical trials. A great deal

of work would need to be done before this technology could be applied in a standardized, cross-culturally sensitive manner to generate primary or secondary measures in clinical trials. That said, the promise of this more objective technique strongly suggests the work should be pursued enthusiastically.

In summary, EMA avoids potential rater inaccuracies and biases associated with patients' recall and provides repeat measurement of a patient's subjective status. Objective data such as location and social context also avoids errors associated with recall over an extensive period. Direct measurement and EMA, used together, may clarify the context and interpretation of findings.

Psychometric Evaluation of Digital Assessment of Negative Symptoms

For remote assessment of negative symptoms to be adopted as primary or secondary endpoints in regulated clinical trials, appropriate psychometric standards will need to be met. Standards for substituting 1 set of measures for another, as well as what constitutes a "gold" reference standard, have not been precisely defined, and a process for defining them has not been developed.[40,41] There is likely no "one-size fits all" solution to psychometric evaluation, and different remote assessments may require very different evaluation approaches. Inter-rater reliability for anhedonia measures, eg, is a critical concern for remote clinical interviews but is of less concern for EMA assessments, given that the latter is inherently subjective and can only be accessed by the reports of a single person. Evaluating the validity of the objective digital phenotyping measures, therefore, raises unique considerations. In evaluating objective biomarkers more generally, some argue that traditionally important aspects of psychometrics are superfluous, given that the intent is to capture tangible (ie, observable and quantifiable) rather than latent phenomena and experienced rather than observed or rated phenomena.[42] Hence, common psychometric metrics, such as internal consistency (ie, evaluating intercorrelations of individual items), and construct and structural validity (ie, measuring the comprehensiveness and structure of measurement) may not apply at all. Ultimately, the psychometric evaluation strategy ie eventually adopted will depend on the clinical inferences being drawn from the measure. Measures meant to tap relatively rare events, eg, getting/being married or getting a job (eg, milestone assessment) will likely require a different plan than for measures meant to tap "in the moment" phenomena or experiences, which will be, by definition, relatively common events.[13] Hence, the "granularity" of the measure, in terms of frequency of assessment and time scale, will be important to consider.

There are general issues to consider with respect to reliability. Perhaps most importantly, the behaviors defining negative symptoms have the potential to be dynamic. Elements of both emotional expression, such as facial expressions, speech rate and prosody, and emotional experiences, such as hedonic experience, can vary within people over minutes, hours, and days and also as they navigate their daily routines.[43,44] The dynamic nature poses a challenge for establishing test-retest reliability, which is targeted at the stability of various behavioral traits. To date, reliability has rarely been reported in studies of remote assessment of negative symptoms, though there is at least some evidence that acceptable to good test-retest reliability can be achieved. EMA and digital phenotyping measures aggregated over blocks of time, eg, by averaging scores on various behaviors and experiences over a time period such as a day or on the basis of some other meaningful feature, such as being at home or away, have resulted in improved test-retest stability (eg, intraclass correlation coefficient values for negative facial expressions from video "selfies" were 0.21 overall vs 0.64 while doing "nothing") which may be acceptable for some purposes.[11–13,24] Beyond test-retest reliability, it is also important to consider systematic changes due to repeated administration (eg, habituation, changes in standards, changes in acuity of self-observation), as these changes could reflect placebo response and/or attenuate sensitivity to treatment effects.

There are general issues with respect to validity to consider. Criterion validity, the degree to which a measure converges with a "gold standard" measure, has been central to validation efforts of remote negative-symptom measures. However, evaluating criterion validity is a challenge in that high convergence may not necessarily be desired given the low yield of studies employing the existing "gold standard" negative-symptom scales. Remote measures were developed, in part, to address potential limitations of traditional measures and hence they differ in many key respects. For example, clinical ratings and EMA measures of anhedonia often show significant but surprisingly modest convergence with each other.[45] When one considers the differences between these measures (eg, that one is based on clinician judgment of symptoms over a 2-week period whereas the other reflects an aggregation of self-reported ratings administered multiple times per day), modest convergence is not surprising. Validity might instead be better understood in the context of whether, for instance, a measure can consistently pick up effects of an intervention. This approach requires another step: establishing the change as clinically significant. One needs to decide whether one wants remote measures to substitute for an in-person interview vs whether one hopes that more objectively quantifiable measures will identify changes not detected with standard clinical scale approaches. For example, biweekly clinician ratings may be biased by tendencies of participants to report today's experience as the average of the inter-visit interval, which could easily miss the variance in experience and activities referenced above.

The repeated and continuously changing nature of remote EMA, with active and passive assessments, offers unique opportunities for evaluating validity (eg, covariance across convergent measures).[21,46] This can be done across many domains of functioning, including those deemed by regulatory agencies as "clinically meaningful."[41,42] For example, several studies have addressed the extent to which participants generally engage in active vs. passive activities. Over time, more physical activity would be expected to relate to better general health and wellness. The association between digital measures of engagement in active vs passive activities and life history variables such as chronicity of illness can be assessed. Measuring change requires understanding of context; eg, in the case of a device measuring geolocation activity, the effects of urbanicity, such as access to transportation, restrictions imposed by a residential setting (eg, locked board and care), and neighborhood crime may be more or as important in activity as one's level of physical health and motivation. Factoring in environmental context is critical to differentiating internally generated and environmentally imposed motivation and energy. This is an important aspect of negative-symptom measure evaluation given the "generalized deficit" issue, concerning the false appearance of specificity due to more global factors that may be immutable by clinical treatments administered in a treatment trial. EMA and devices measuring geolocation activity supplement information from the interview because they bypass the limitations of potentially cognitively impaired patients to recall details of their activities and the specifics of their location over a lengthy assessment period.

Another general consideration in developing an evaluation strategy involves understanding systematic influences on reliability and validity from demographic, cultural, linguistic, and other individual differences. Systematic differences in clinically rated negative symptoms have been reported, eg, between men and women and Black and White patients.[47,48] Complicating this is the reality that the behaviors underlying negative symptoms can differ dramatically as a function of demographic factors and even individual life histories. The meaning of a smile, eg, and its appropriateness in expressing it while interacting with a medical doctor, has been shown to differ between European and Asian cultures.[49] Potential biases can be exacerbated with digital phenotyping measures that rely on computerized algorithms for processing and interpreting objective data if not derived from demographically diverse samples. An algorithm defining smiles developed on Europeans, eg, may be far from optimal in being applied to Asian cultures. Awareness of potential "biases" in big data applications is increasing, as are calls for government oversight to try to address them. Developing culturally sensitive measures of negative symptoms is critical for addressing issues of systemic racism and inequality in psychosis assessment/treatment.[50]

Standardization of Digital Assessment of Negative Symptoms

Remote assessment introduces new challenges for standardization and normative understanding and it is not yet clear how to manage this. Measures can vary, eg, in timing, frequency, and nature of how data are collected, and this could lead to profound differences in their potential scores. Measures of hedonic experience often systematically vary as a function of diurnal cycle as do activities such as resting vs. watching television, so data collection strategies, including mid-day measures, will likely vary from those that do not.[51] Certain activities that are normative for healthy people at certain times of day (eg, seated during mid-day and watching television in the evening) may not be normative if they are engaged in all day. For example, being home, alone, and seated for 6 consecutive hours may be highly adaptive for software developers, but a sign of avolition or asociality in an individual with serious mental illness. Normative standards for social and recreational activities are completely lacking, and, as noted above, behaviors that may reflect maladaptive activities in some participant groups may be central to adaptive success in others.

Beyond data collection, data processing approaches vary and may be proprietary; this is a notable challenge for understanding voluminous and high-dimensional data. Complex objective data are dependent on extensive processing to extract features and derive summary values for interpretation, and this process involves many decisions that will be difficult to standardize across studies. Standardizing and achieving transparency with respect to these decisions will be particular challenges for studies using machine learning and artificial intelligence. The infusion of private sector solutions, which may contain proprietary solutions and intellectual property, can further complicate this process. Finally, accounting for differences in operating systems, screen size, recording technology, connectivity speed, and device processing speed and their evolution over time will be important. Relatedly, there is a lack of standardization in regulatory and privacy standards across countries, and it is likely that these will evolve over time.

Conclusions and Next Steps

A number of factors led to an increased interest in the remote assessment of negative symptoms. The most obvious was the COVID pandemic, which resulted in investigators and subjects being more comfortable with the tools for the remote use of rating instruments that were developed for in-person rating. At the same time, advances in the development of mobile digital devices and methods for analyzing large data sets also suggested new approaches to measuring negative symptoms. Finally, investigators realized that there is a distinct limitation of rating negative symptoms in a clinic; ie, it depends on a subject's ability to recall their

interests and motivations in the past. Together, these factors result in a clear and urgent need for reliable, valid, generalizable, and safe remote measures of negative symptoms. Beyond substituting for in-clinic measures, remote measures have the potential of identifying clinically meaningful phenomena that are not as sensitively detected in clinical interviews. Despite over 4 decades of progress toward this goal, clinical trials lack any secondary or primary outcome measures that can be implemented at this time.

The pandemic led to a rapid transition from in person to remote administration of negative symptoms rating instruments before the remote methods were validated. It is important that this validation compare ratings of the same subjects in person and remotely. There are other challenges for evaluating the validity of passive and active remote assessment using digital devices, objective measures of expressiveness, and laboratory-based objective measures of motivation. These may be measuring something more closely aligned with negative-symptom psychopathology than the so-called gold standard clinical rating scales. An NIMH-MATRICS-like project would likely facilitate consensus on psychometric evaluation and standardization procedures and hasten adoption of these measures. Beyond psychometric proprieties, the rate and the depth of an assessment scale or assessment technologies adoption also depend on need, ease of use, and cost. There are unplanned circumstances that accelerate adoption. The fact that PANSS, eg, was utilized in the pivotal trials of risperidone, the first antipsychotic to be marketed after a long period, established the scale as the main instrument to assess severity of symptoms in schizophrenia. A similar disruptive event could accelerate the adoption of remote assessment technologies for negative symptoms.

Acknowledgments

Dr Raeanne C. Moore is a co-founder of KeyWise AI, Inc. and a consultant for NeuroUX. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. Dr Stephen R. Marder holds the position of Co-Editor-in-Chief for Schizophrenia Bulletin Open and has been recused from reviewing or making decisions for the manuscript.

References

1. Marder SR, Galderisi S. The current conceptualization of negative symptoms in schizophrenia. *World Psychiatry.* 2017;16(1):14–24.
2. Axelrod BN, Goldman RS, Alphs LD. Validation of the 16-item Negative Symptom Assessment. *J Psychiatr Res.* 1993;27(3):253–258.
3. Daniel D, Alphs L, Cazorla P, Bartko J, 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(2):87–94.
4. Kirkpatrick B, Strauss GP, Nguyen L, *et al.* The brief negative symptom scale: psychometric properties. *Schizophr Bull.* 2010;37(2):300–305.
5. Kring AM, Gur RE, Blanchard JJ, Horan WP, Reise SP. The Clinical Assessment Interview for Negative Symptoms (CAINS): final development and validation. *Am J Psychiatry.* 2013;170(2):165–172.
6. Edgar CJ, Blaettler T, Bugarski-Kirola D, Le Scouiller S, Garibaldi GM, Marder SR. Reliability, validity and ability to detect change of the PANSS negative symptom factor score in outpatients with schizophrenia on select antipsychotics and with prominent negative or disorganized thought symptoms. *Psychiatry Res.* 2014;218(1–2):219–224.
7. Sharp IR, Kobak KA, Osman DA. The use of videoconferencing with patients with psychosis: a review of the literature. *Ann Gen Psychiatry.* 2011;10(1):10–14.
8. Zarate CA, Weinstock L, Cukor P, *et al.* Applicability of telemedicine for assessing patients with schizophrenia. *J Clin Psychiatry.* 1997;58(1):22–25.
9. Geddes MR, O'Connell ME, Fisk JD, Gauthier S, Camicioli R, Ismail Z. Remote cognitive and behavioral assessment: report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. *Alzheimers Dement.* 2020;12(1) :1–11.
10. Kobak KA, Williams JBW, Engelhardt N. A comparison of face-to-face and remote assessment of inter-rater reliability on the Hamilton Depression Rating Scale via videoconferencing. *Psychiatry Res.* 2008;158(1):99–103.
11. Cohen AS, Cowan T, Le TP, *et al.* Ambulatory digital phenotyping of blunted affect and alogia using objective facial and vocal analysis: proof of concept. *Schizophr Res.* 2020;220:141–146.
12. Cohen AS, Schwartz E, Le TP, *et al.* Digital phenotyping of negative symptoms: the relationship to clinician ratings. *Schizophr Bull.* 2020;47(1):44–53.
13. Granholm E, Holden JL, Mikhael T, *et al.* What do people with schizophrenia do all day? Ecological momentary assessment of real-world functioning in schizophrenia. *Schizophr Bull.* 2020;46:242–251. doi:10.1093/schbul/sbz070
14. Parrish EM, Depp CA, Moore RC, *et al.* Emotional determinants of life-space through GPS and ecological momentary assessment in schizophrenia: what gets people out of the house? *Schizophr Res.* 2020;224:67–73.
15. Moran EK, Culbreth AJ, Barch DM. Ecological momentary assessment of negative symptoms in schizophrenia: relationships to effort-based decision making and reinforcement learning. *J Abnorm Psychol.* 2017;126(1):96–105.
16. Ben-Zeev D. Mobile technologies in the study, assessment, and treatment of schizophrenia. *Schizophr Bull.* 2012;38(3):384–385.
17. Jones SE, Moore RC, Pinkham AE, Depp CA, Granholm E, Harvey PD. A cross-diagnostic study of adherence to ecological momentary assessment: comparisons across study length and daily survey frequency find that early adherence is a potent predictor of study-long adherence. *Per Med Psychiatry.* 2021;29–30:100085.
18. Strauss GP, Horan WP, Kirkpatrick B, *et al.* Deconstructing negative symptoms of schizophrenia: avolition–apathy and diminished expression clusters predict clinical presentation and functional outcome. *J Psychiatr Res.* 2013;47(6):783–790.
19. Strauss GP, Zamani Esfahlani F, Sayama H, *et al.* Network analysis indicates that avolition is the most central domain for

the successful treatment of negative symptoms: evidence from the ROLUPERIDONE randomized clinical trial. *Schizophr Bull.* 2020;46(4):964–970.

20. Durand D, Strassnig MT, Moore RC, *et al.* Self-reported social functioning and social cognition in schizophrenia and bipolar disorder: using ecological momentary assessment to identify the origin of bias. *Schizophr Res.* 2021;230:17–23.
21. 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(1–3):24–30. https://pubmed.ncbi.nlm.nih.gov/34150360/. Accessed October 23, 2022.
22. Strassnig MT, Harvey PD, Miller ML, Depp CA, Granholm E. Real world sedentary behavior and activity levels in patients with schizophrenia and controls: an Ecological Momentary Assessment Study. *Ment Health Phys Act.* 2021;20:100364.
23. Strassnig MT, Miller ML, Moore R, Depp CA, Pinkham AE, Harvey PD. Evidence for avolition in bipolar disorder? A 30-day ecological momentary assessment comparison of daily activities in bipolar disorder and schizophrenia. *Psychiatry Res.* 2021;300:113924.
24. Raugh IM, James SH, Gonzalez CM, *et al.* Geolocation as a digital phenotyping measure of negative symptoms and functional outcome. *Schizophr Bull.* 2020;46(6):1596–1607.
25. Depp CA, Bashem J, Moore RC, *et al.* GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. *NPJ Digit Med.* 2019;2(1):1–7. doi:10.1038/s41746-019-0182-1
26. Jones SE, Moore RC, Depp CA, Ackerman RA, Pinkham AE, Harvey PD. Daily ecological momentary assessments of happy and sad moods in people with schizophrenia and bipolar disorders: what do participants who are never sad think about their activities and abilities? *Schizophr Res: Cogn.* 2021;26:100202.
27. Brenner CJ, Ben-Zeev D. Affective forecasting in schizophrenia: comparing predictions to real-time ecological momentary assessment (EMA) ratings. *Psychiatr Rehabil J.* 2014;37(4):316–320.
28. Gard DE, Kring AM, Gard MG, Horan WP, Green MF. Anhedonia in schizophrenia: distinctions between anticipatory and consummatory pleasure. *Schizophr Res.* 2007;93(1–3):253–260.
29. Cohen AS, Minor KS. Emotional experience in patients with schizophrenia revisited: meta-analysis of laboratory studies. *Schizophr Bull.* 2010;36(1):143–150.
30. Kring AM, Barch DM. The motivation and pleasure dimension of negative symptoms: neural substrates and behavioral outputs. *Eur Neuropsychopharmacol.* 2014;24(5):725–736.
31. Visser KF, Chapman HC, Ruiz I, Raugh IM, Strauss GP. A meta-analysis of self-reported anticipatory and consummatory pleasure in the schizophrenia-spectrum. *J Psychiatr Res.* 2020;121:68–81.
32. Cohen AS, Elvevåg B. Automated computerized analysis of speech in psychiatric disorders. *Curr Opin Psychiatry.* 2014;27(3):203–209.
33. Meehl PE. Schizotaxia, schizotypy, schizophrenia. *Am Psychol.* 1962;17(12):827–838.
34. Sullivan H. *Schizophrenia as a Human Process*. New York: W. W. Norton and Company; 1962.
35. Kirkpatrick B, Mucci A, Galderisi S. Primary, enduring negative symptoms: an update on research. *Schizophr Bull.* 2017;43(4):730–736.
36. Chan RCK, Wang Y, Huang J, *et al.* Anticipatory and consummatory components of the experience of pleasure in schizophrenia: cross-cultural validation and extension. *Psychiatry Res.* 2010;175(1–2):181–183.
37. Buck B, Minor KS, Lysaker PH. Lexical characteristics of anticipatory and consummatory anhedonia in schizophrenia: a study of language in spontaneous life narratives. *J Clin Psychol.* 2015;71(7):696–706.
38. Cohen AS, St-Hilaire A, Aakre JM, Docherty NM. Understanding anhedonia in schizophrenia through lexical analysis of natural speech. *Cogn Emot.* 2009;23(3):569–586.
39. Horan WP, Reddy LF, Barch MF, *et al.* Effort-based decision-making paradigms for clinical trials in Schizophrenia: Part 2—external validity and correlates. *Schizophr Bull.* 2015;41(5):1055 –1065. doi: 10.1093/schbul/sbv090
40. Reddy LF, Horan WP, Barch DM, *et al.* Effort-based decision-making paradigms for clinical trials in schizophrenia: part 1—psychometric characteristics of 5 paradigms. *Schizophr Bull.* 2015;41(5):1045–1054.
41. Center for Drug Evaluation and Research. *Digital Health Technologies for Remote Data Acquisition in Clinical Investigation.* U.S. Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations. Accessed October 23, 2022.
42. Center for Drug Evaluation and Research. *Patient-Reported Outcome Measures: Use in Medical Product Development*. U.S. Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims. Accessed October 23, 2022.
43. Cohen AS, Fedechko TL, Schwartz EK, *et al.* Ambulatory vocal acoustics, temporal dynamics, and serious mental illness. *J Abnorm Psychol.* 2019;128(2):97–105.
44. Wright AG, Hopwood CJ. Advancing the assessment of dynamic psychological processes. *Assessment.* 2016;23(4):399–403.
45. Moran EK, Culbreth AJ, Barch DM. Emotion regulation predicts everyday emotion experience and social function in schizophrenia. *Clin Psychol Sci.* 2017;6(2):271–279.
46. Harvey CD, Moore RC, Depp CA, Ackerman RA, Pinkham AE, Harvey PD. The Association of Momentary Sad Moods, concurrent productive behavior, and global functional outcomes: a 30-day ecological momentary assessment study of people with bipolar illness. *Cognit Neuropsychiatry.* 2022;27(5):342–355.
47. Ahmed AO, Kirkpatrick B, Galderisi S, *et al.* Cross-cultural validation of the 5-factor structure of negative symptoms in schizophrenia. *Schizophr Bull.* 2018;45(2):305–314.
48. Trierweiler SJ, Neighbors HW, Munday C, Thompson EE, Binion VJ, Gomez JP. Clinician attributions associated with the diagnosis of schizophrenia in African American and non-African American patients. *J Consult Clin Psychol.* 2000;68(1):171–175.
49. Matusitz J, Spear J. Doctor-patient communication styles: a comparison between the United States and three Asian countries. *J Hum Behav Soc Environ.* 2015;25(8):871–884.
50. Olbert CM, Nagendra A, Buck B. Meta-analysis of black vs. white racial disparity in schizophrenia diagnosis in the United States: do structured assessments attenuate racial disparities? *J Abnorm Psychol.* 2018;127(1):104–115.
51. Bylsma LM, Rottenberg J. Uncovering the dynamics of emotion regulation and dysfunction in daily life with ecological momentary assessment. In: Nykliček I, Vingerhoets A, Zeelenberg M, eds. *Emotion Regulation and Well-Being*. Springer Science + Business Media, 2011:225–244. doi:10.1007/978-1-4419-6953-8_14