

PRIMARY RESEARCH

Open Access

Site-independent confirmation of subject selection for CNS trials: 'dual' review using audio-digital recordings

Steven D Targum* and J Cara Pendergrass

Abstract

Background: *Site-independent* review of subject eligibility for central nervous system (CNS) trials has been used as a surveillance method to enhance the integrity and precision of the subject selection process. We evaluated the utility of a customized review strategy that employs site-independent review of audio-digital recordings of site-based screen interviews.

Methods: We applied a customized site-independent subject selection strategy in nine phase II double-blind, placebo-controlled clinical trials across the CNS spectrum. The Clinical Validation Inventory for Study Admission (C-VISA™, Boston, MA, USA) was developed as a site-independent review method that evaluates and confirms diagnoses, symptom severity, and subject validity prior to enrollment (randomization) into a clinical trial. The C-VISA™ method uses audio-digital recordings of actual site-based interviews conducted at the screening visit. The recordings of these interviews accompanied by digital notes are electronically submitted for independent review and 'dual' scoring of key rating instruments. A multi-tiered system of site-independent reviewers either affirms subject eligibility or identifies administrative and/or clinical issues that may preclude study eligibility (screen failure).

Results: In this meta-analysis, 404 of 2,515 submitted C-VISA™ eligibility reviews (16.1%) were challenged by tier 1 reviewers and escalated to a tier 2 reviewer. After telephone adjudication with the respective trial site investigator, 168 of these 404 tier 2 reviews (41.6%) were not approved yielding an overall screen fail rate of 6.7% for all C-VISA™ submissions. The primary reasons for screen failure were insufficient documentation to support the intended diagnosis, symptom severity that did not meet protocol criteria, the presence of excluded comorbid conditions, and potential confounding factors that might obscure assessment during the trial.

Conclusion: The C-VISA™ review process coupled with dual independent scoring of key rating instruments is a quality assurance strategy that provides a systematic site-independent eligibility filter to enhance the precision of subject selection and the integrity of study data. The C-VISA™ strategy has broad applicability across the CNS spectrum because it achieves the objective of confirmatory site-independent review without producing excessive site or subject burden.

Keywords: Subject validity, Subject selection, Clinical trials, Drug development, Surveillance: audio-digital recording, Site-independent review, Rater competency

* Correspondence: sdtargum@yahoo.com
Clintara LLC, 11 Beacon Street, Suite 600, Boston, MA 02108, USA

Background

Central nervous system (CNS) drug development has been challenged by the frequent failure to achieve signal detection in many clinical trials [1-7]. Numerous factors affect trial outcomes including the effectiveness of the candidate drug, study design, study completion, and trial execution [6-8]. Two important trial execution factors are subject selection and ratings precision [6-12]. Inappropriate subject selection and inaccurate ratings reduce the statistical power to achieve signal detection [13-15].

Appropriate subject selection is a complex process that requires collection of accurate documentation and reliable symptomatic measurements at a time when the potential study subject is often least well known to the site (screening) and often most disturbed (ill). Potential study candidates may be unfamiliar with study procedures and arbitrarily endorse symptoms or exaggerate severity in order to gain entry into the clinical trial. Misplaced site-based incentives (e.g., financial, humanitarian, and competitive interests) may also influence subject selection because clinical trial sites are generally incentivized to enroll rather than reject subjects [6,7,16-18]. Further, there may be concurrent, confounding factors or comorbid conditions that might adversely affect the reliability of symptom severity ratings during the course of the clinical trial [11]. For instance, accurate assessments might be obscured in study candidates who have experienced recent trauma or other de-stabilizing environmental conditions (e.g., unstable housing, family discord, recent job loss). Consequently, subject selection for CNS trials is a complex, risky process that requires more careful attention and scrutiny.

The inherent challenges of subject selection have generated a compelling argument to use a 'dual' eligibility strategy that combines *site-independent* reviews as unbiased 'second' opinions in conjunction with the site-based assessments. Site-independent reviews attempt to manage the potential for insufficient documentation, incomplete interviews, inaccurate diagnoses, or inflated screen scores [8,11,18-20]. These reviews are quality assurance surveillance 'filters' that improve the integrity and precision of subject selection. It has been shown that assessment and reliability improve when more than one rater assesses the same subject and provides a second opinion [14,21]. The mere awareness of a *secondary* review (surveillance) can often improve the quality of the initial site-based assessment as well.

Site-independent clinical assessments by live telephone, audio, or video interviews have been used across the CNS spectrum for many years [8,22-26]. These external assessments increase the time and the cost of doing a study and can add considerable site-burden and resistance as well. Consequently, an external eligibility review strategy must be reasonably balanced between the objective of enhancing subject selection

and the need to sustain the engagement of the subject and collaboration of the trial site.

A subject who is truly appropriate ('valid') for enrollment in a clinical trial must meet additional selection criteria that exceed the conventional protocol entry criteria of most CNS trials such as excluding subjects with potentially confounding (de-stabilizing) conditions that may obscure assessment [11,27,28]. Most surveillance strategies do not assess the appropriateness ('validity') of subjects for participation in a clinical trial. The SAFER criteria inventory was introduced as an external review strategy to affirm validity of subjects for CNS clinical trials [11,19]. The SAFER assessment is administered as a telephone interview by a site-independent clinician between the screen and baseline visit and requires extra site and subject time for scheduling and coordination [11,19]. As an entirely separate interview, the SAFER interview is subject to informational and temporal variance because the remote SAFER telephone interviewer may obtain different responses from a subject who may be more or less informative relative to the preceding 'live' site-based assessment.

An alternative method to obtain a second opinion about subject eligibility is to record and independently assess the actual screen interviews conducted by the site-based rater. For instance, Schoemaker and colleagues used laptop computers with inbuilt cameras to record some site-based assessments [20]. In their system, recorded interviews were transmitted over the web via a secure server for site-independent review of symptom severity [20]. This method eliminated both temporal and informational variance between site-based and independent raters because both raters evaluated exactly the same information.

This paper describes the application of a recording strategy for site-independent review. The Clinical Validation Inventory for Study Admission (C-VISATM, Houston, TX, USA) uses audio-digital recordings of key site-based screening assessments as the basis for site-independent subject eligibility review. The recordings facilitate a dual screening assessment by using site-independent clinicians to review the actual site-based screening interviews. This external review strategy was developed to enhance the reliability of subject selection for CNS trials with a procedure that could be applied in large global clinical trials and would not require extra time or site burden.

The aim of this paper was to examine the utility of the C-VISATM dual review eligibility process in nine double-blind, placebo-controlled clinical trials conducted across the CNS spectrum including studies of subjects with major depressive disorder (MDD), schizophrenia, and Alzheimer's disease (AD).

Methods

Audio-digital recordings and customized C-VISATM screening assessment batteries were used as part of the

subject selection process in nine phase II double-blind, placebo-controlled clinical trials conducted across the CNS spectrum. The data from these studies were merged for this meta-analysis and include two studies of patients with MDD, five studies in subjects with schizophrenia, and two studies of mild-moderate Alzheimer's disease. Clinical trial sites from the United States, Canada, Europe, and Asia participated in these studies. All potential study subjects signed a written consent approved by an institutional review board that included consent to audio-digital recording of key screen assessments using a customized C-VISA™ workbook as part of the screening process. A trained, site-based rater administered the C-VISA™ assessment battery (described below) using a commercially available audio-digital pen. The recordings were electronically transmitted via a secure website to Clintara LLC (Boston, MA, USA) who coordinated a multi-layered site-independent review process. The external review process included both administrative review for completeness and two tiers of clinical review to evaluate subject eligibility.

Development of the C-VISA™

The C-VISA™ contains study-specific screening instruments and a customized worksheet developed for investigators to summarize the clinical entry criteria supporting subject eligibility. This documentation may include evidence gathered from multiple sources. The C-VISA™ was derived, in part, from the principles used to develop the SAFER validity criteria and the concept of validity as described by Kendell and Jablensky to ascertain whether a subject's clinical presentation and recent history is 'rare' enough (valid) to be appropriate for a specific clinical trial [11,27].

A customized C-VISA™ workbook was developed for each of the nine studies in this analysis based upon protocol-specific criteria. Each C-VISA™ included a battery of assessment components:

- a) A diagnostic screening module that was either a historical narrative or validated screening instrument such as the Mini International Neuropsychiatric interview (M.I.N.I.) or Structured Clinical Interview for DSM-IV, Structured Clinical Interview for DSM diagnoses (SCID) [29,30]
- b) A clinician-administered symptom severity questionnaire identifying and quantifying current clinical symptoms related to the symptom severity thresholds as delineated in the protocol
- c) A narrative impression of global illness severity that documented the impact of current symptoms on behavior and function [12]
- d) A C-VISA™ summary worksheet customized to address all inclusion and exclusion criteria and

explore potential confounding (or destabilizing) factors and reliability issues that might obscure assessment [31]

In each study, the customized C-VISA™ summary worksheet required documentation to affirm that the subject was truly appropriate (valid) for the specific study and met the following validity criteria:

1. *Diagnostic verification.* A clinical presentation that met diagnostic criteria and had face validity consistent with the typical course of the underlying disease state
2. *Symptom severity confirmation.* Sufficient acute symptom severity that scored within the entry criteria thresholds established by the protocol
3. *Reliable and measurable symptoms.* Sufficient, *measurable* symptoms based upon the selected rating instruments that might be sensitive to change during the specified time interval and could be reliably assessed with the subject (and informant as needed) during the clinical trial
4. *Absence of confounding factors.* Subject did *not* have any co-morbid conditions, excluded medications, or confounding factors that might obscure meaningful measurement of clinical change during the study (e.g., substance abuse, recent trauma)
5. *Clinical relevance.* Sufficient documented evidence that the current symptoms were clinically relevant *and* warranted treatment intervention

Clinical relevance, as noted above, meant that the presenting symptoms had some pathological impact and were therefore *consequential* to the subject's behavior or function. In studies involving psychosis or cognitive impairment, the documentation of clinical relevance required collateral information and/or corroboration from a reliable informant.

Audio-digital recordings

Customized C-VISA™ workbooks were designed to work in conjunction with audio-digital pen recorders. These writing instruments were commercially available pens that have both audio recording and digital image capture technologies. The customized C-VISA™ workbooks were manufactured on specialized paper forms that capture the digital images (written notes) which were uploaded with the audio recordings and electronically transmitted via a secure website to Clintara LLC for independent review.

Site-independent review process: implementation of C-VISA™ in a clinical trial

In all nine studies, the site-based raters obtained the subject's consent to record the screening interviews

using the audio-digital pen recorder prior to commencing the interviews. As described above, the site-based rater notes were graphically digitized on specialized paper and accompanied the audio recordings obtained with the recording pen. In addition, the trial site investigator (or designee) reviewed all of the available screening information and completed the comprehensive C-VISA™ summary worksheet that reviewed and examined the fundamental issues related to subject eligibility and validity.

The completed C-VISA™ battery was uploaded via a secure website to a central site (Clintara LLC) for distribution for site-independent review. The site-independent review process included an initial administrative review for data completeness and a two-tiered clinical review. Site-independent clinical reviewers were subcontracted regional experts who were trained in the study procedures and spoke the language of the raters. The site-independent reviewers had extensive experience in clinical trials with the designated study population and completed the same rater-training program and met the same qualification standards to conduct ratings as the site-based raters. Clintara employed an internal quality control program that included a random, secondary independent review of the primary site-independent reviewers (an independent review of the reviewers) to affirm the sustained reliability of their assessments.

Site-independent reviewers were not affiliated with the clinical trial site submitting the proposed study candidate for the study. Site queries could occur at any point during this layered review process to request further documentation or clarification about the submitted screen materials or clinical queries. Clinical review included a primary clinical reviewer (tier 1) who reviewed every subject submission and a second clinical reviewer (tier 2) whenever there was a tier 1 eligibility challenge. The tier 2 review was an interactive process that included adjudication (telephone discussion) with the site investigator as part of a collaborative and consultative process that was considered an essential part of

the review process. The study charter authorized the site-independent reviewer to make the final recommendation regarding subject eligibility *after* telephonic adjudication with the trial site investigator. The timeline objective was to complete the site-independent review process within 72 h of receipt of materials in order to facilitate the randomization of appropriate subjects.

Results

Subject selection across all studies

Each of the nine studies included in this meta-analysis experienced a high rate of administrative review when the studies were initiated. Across the nine studies, the rates of administrative review ranged from 50%–90% of submissions within the first 6 months. Most of the administrative issues related to insufficient documentation could be corrected and thereby avoid later queries. In all studies, the quality of interviews and submitted screen data improved as a consequence of external review and remediation. Consequently, the rate of administrative reviews for incomplete documentation dropped considerably as the study proceeded. The rates of administrative review ranged from 25%–40% of submitted C-VISA™ batteries *after* the first 6 study months.

For this analysis, 2515 C-VISA™ batteries passed administrative review and were submitted for independent clinical review. The mean time for final determination of study eligibility across this entire sample of reviews was 55.8 ± 24.5 h (SD). The mean turnaround time included the adjudication process with the trial site investigator. The addition of the audio-digital recording requirement and external review procedure did not affect the rate of enrollment in these studies.

As shown in Table 1, tier 1 clinical reviewers challenged 404 of the C-VISA™ reviews (16.1% of all submissions). One hundred sixty-eight of these challenged reviews (41.6%) were ultimately screen failed by tier 2 reviewers after telephone adjudication with the trial site investigator.

Table 1 Subject selection issues identified across the CNS spectrum

All reviewed subjects	Challenged cases (% of all cases reviewed)		Not approved (% of all challenged cases)	
Inadequate documentation to support the diagnosis	68	2.7%	19	27.9%
Insufficient symptom severity (too low)	75	2.9%	41	54.7%
Symptom severity too high (above entry criteria)	44	1.8%	28	63.6%
Presence of comorbid medical diagnosis	30	1.2%	5	16.6%
Presence of comorbid psychiatric diagnosis	67	2.7%	29	43.3%
Substance abuse/dependence	29	1.2%	12	41.4%
External confounding factors	50	1.9%	11	22.0%
Unreliable/unassessable subject or informant	13	0.5%	6	46.2%
Other protocol-specific study criteria	28	1.1%	17	60.7%
Total of <i>all</i> submitted subjects ($n = 2,515$)	404	16.1%	168	6.7%

Thus, the overall C-VISATM screen fail rate across all studies was 6.7% of all submitted reviews.

As noted in Table 1, the most common reasons for tier 2 escalations across all studies included inadequate documentation to support the intended diagnosis, symptom severity scores that did not fall within the protocol criteria thresholds (either too high or too low), and the presence of excluded comorbid medical or psychiatric conditions. Potential confounding (or destabilizing) factors that might obscure assessment of symptom severity during the trial were another common reason for challenging subject eligibility, particularly in the schizophrenia studies. Confounding factors included recent exposure to traumatic events, personal losses, unstable living conditions, or relocation to a new area without social supports. In addition, the independent reviewers identified some inadequate interviews that led to rater remediation or circumstances in which the informant (often the caregiver for AD studies) was not reliable.

Sixty-eight cases (2.7%) were initially challenged because of inadequate documentation to support the diagnosis. Clinical queries asked for more information about the onset and behavioral impact of symptoms, documentation about current symptoms to meet diagnostic criteria, or evidence of progressive decline (in Alzheimer's disease studies). In more than 72% of cases, sufficient documentation was provided to support the diagnosis and approve the case. In contrast, only 42% of cases challenged regarding symptom severity criteria were ultimately approved (see Table 1). The reasons for tier 2 escalations varied somewhat by disease category and the study-specific entry criteria, but the rates of challenge and approval were largely consistent across the CNS spectrum.

Subject selection for depression studies

In the depression studies, the C-VISATM batteries included review of the recorded diagnostic screening instrument and dual scoring of the symptomatic severity eligibility measurement. Excluded psychiatric diagnoses (bipolar disorder or substance abuse) were identified in 2.5% of reviewed cases as part of the diagnostic assessments but had been overlooked or minimized as relevant by the site investigator prior to the tier 2 discussion. Further, dual scoring of the recorded screen symptomatic questionnaires identified 5.9% of subjects who were below the minimum severity threshold for study entry (insufficient symptom severity). Once informed via remediation, these putative cases of score inflation were rarely repeated on subsequent assessments by the same site-based rater.

Subject selection for schizophrenia studies

Data was analyzed from five separate studies of patients with schizophrenia. In our analysis, 3.5% of these patients presented with concomitant mood symptoms

(and the possibility of schizoaffective disorder) that were study exclusions and therefore challenged by tier 1 reviewers. Overall, tier 2 reviewers screen failed 2% of all submissions for the schizophrenia studies because of excluded comorbid psychiatric conditions.

These studies had a range of acceptable symptom severity eligibility scores based upon the Brief Psychiatric Rating Scale (BPRS) or Positive and Negative Syndrome Scale (PANSS) total screen scores [32,33]. Four studies sought patients with an acute exacerbation of psychosis (with a minimum symptom severity threshold) whereas one study sought patients with relatively stable illness (with a maximum illness severity threshold). Dual site-independent scores were based upon the audio recordings of the BPRS or PANSS interviews with accompanying corroborative digital notes. Dual scores resulted in screen failures for 1.8% of patients because of insufficient symptom severity and 1.0% because of excessive severity exceeding protocol criteria thresholds.

In one study, we examined the trial outcome of 39 'challenged' subjects who achieved tier 2 approval after adjudication and proceeded to randomization (DeMartinis, personal communication). The completion rate and trial outcome of these subjects in this study was similar to the other 273 subjects who were unchallenged by the tier 1 reviewers; 76% of all randomized subjects completed the study, and 74% of randomized tier 2 approvals completed the study. Further, there were no substantive differences between tier 1 and tier 2 approved subjects on baseline PANSS total score or any end point scores.

Subject selection for Alzheimer's disease studies

Clinical trials of AD rely on historical documentation for progressive cognitive decline to make the diagnosis of probable AD and do not use a standardized diagnostic screening instrument like the M.I.N.I. or SCID. Therefore, the customized C-VISATM batteries for the AD studies included a narrative historical section to support and verify the diagnosis. In these studies, the additional modules included the Clinical Dementia Rating Scale (CDR), Alzheimer's Disease Assessment Scale-cognitive part (ADAS-cog), and/or Mini-Mental State Examination to allow for dual independent scoring of the site-based recordings [34-36].

In the AD studies, insufficient historical documentation of progressive cognitive decline was often an administrative issue that could easily be resolved following queries to the trial site. Symptom severity scores that were too low or too high and did not meet the protocol entry criteria (as measured by dual scoring of the Mini-Mental State Exam or ADAS-cog) accounted for the majority of screen failures in the AD studies. The presence of comorbid medical or psychiatric conditions (e.g., unstable medical

illness, agitation, or psychosis) and possible confounding factors (living conditions, caregiver reliability) were also frequently identified for tier 2 reviews.

Discussion

This paper describes the application of a novel site-independent review procedure to affirm subject eligibility for CNS trials. The Clinical Validation Inventory for Study Admission (C-VISA™) uses audio-digital recordings of key site-based screening assessments as the basis for external (site-independent) eligibility determination. We have developed a scalable external review procedure that can improve subject selection and ratings precision for CNS trials, can be applied in global clinical trials regardless of language or geography, and does not add subject or site burden.

As described in this paper, we applied the C-VISA™ review strategy in nine double blind, placebo-controlled clinical trials across the CNS spectrum. In this meta-analysis, 16.1% of all 2515 C-VISA™ submissions were challenged by tier 1 reviewers and escalated for further review. Tier 2 telephone adjudication with the site investigator clarified missing information or resolved clinical queries and facilitated approval of more than half (58.4%) of the challenged cases. In fact, more than 72% of cases challenged for lack of sufficient documentation were approved after adjudication. Overall, site-independent reviewers screen failed 6.7% of all submitted C-VISA™ submissions. Some investigators acknowledged privately that they increased their own level of eligibility diligence and were reluctant to submit questionable subjects because of the rigor of the external review process. Therefore, the mere presence of surveillance may have abetted some inappropriate submissions.

In all studies, the quality of submitted screen data improved as a consequence of external review, rater remediation, and adjudication. The rate of administrative reviews dropped as trial sites recognized that the administrative reviewers were actually identifying a lack of sufficient documentation. Therefore, we believe that the ongoing surveillance process improved vigilance at the trial sites, improved documentation, reduced later queries, and contributed to better subject selection.

The audio-digital recording method was scalable and able to manage the complexity of global trial sites in Europe and Asia as well as the United States. Site-independent reviewers were selected from those regions and were familiar with the language and cultural issues present in these populations. The process did not disrupt clinical or operational procedures at the trial sites. The data was easily transmitted via a secure website to the external reviewers, was reviewed quickly without delaying randomization of appropriate subjects, and did not slow down the sponsor's enrollment timelines.

The use of the C-VISA™ surveillance strategy for subject selection

The C-VISA™ workbook includes a battery of key study eligibility assessments and a summary worksheet that is customized for each study based upon the protocol-specific requirements. Generally, the screening battery incorporates a standardized diagnostic interview instrument, validated symptomatic measurement scales that identify the presence and severity of the targeted symptoms of interest (usually aligned to protocol specific symptom severity thresholds), a global or functional assessment measure that requires written documentation about the presence and clinical relevance of the 'targeted' symptoms of interest, and other historical documentation or treatment history forms to ascertain the appropriateness (validity) of the subject for this specific study.

The external review takes place prior to the baseline visit (usually randomization) by a *site-independent* clinical expert who assesses exactly the same clinical material obtained and recorded at the site by the site-based interviewer. This procedure generates a second, site-independent subject eligibility review that can effectively derail any potential site-based misincentives that might bias enrollment decisions. The idea that two separate assessments of the same subject can improve reliability and reduce methodological errors in clinical trials is not a new concept and has been applied by other investigators [13,14]. A dual affirmation strategy improves subject selection and may optimize trial outcomes [18,31,37].

External review provides an unbiased diagnostic verification and symptom severity confirmation to support subject eligibility. External review is a quality assurance filter to affirm that the site assessments were appropriately completed according to the established protocol. Beyond this obvious surveillance function, the independent reviewers may identify potentially confounding factors that could contribute to fluctuating symptom severity and affect the trial outcome. The potential impact of confounding factors may be overlooked or underestimated by the site-based rater assessing the subject for enrollment. For instance, recent traumatic experiences, geographic relocations, unstable living conditions, personal losses, or family crises may be non-specific factors that reduce the sensitivity of the symptomatic measurements to the experimental drug treatment. Consequently, confounding factors could impede detection of a significant drug treatment effect during the clinical trial. A site-independent, unbiased reviewer is more likely to focus on these potentially confounding factors or validity issues than a busy site-based clinician who is primarily focused on enrollment.

The challenges of subject selection

The three major reasons for screen failure across the CNS studies in this meta-analysis (regardless of diagnostic

category) were insufficient documentation to support the diagnosis, excluded comorbid psychiatric conditions, and symptom severity scores that did not meet the protocol entry criteria thresholds (scores that were either too low or too high). We have identified several issues that often lead to eligibility challenges, require adjudication with the investigator and may result in screen failure:

1. Inadequate documentation to verify the diagnostic criteria
2. Insufficient or excessive symptom severity (by protocol-specific criteria)
3. Excluded comorbid medical/psychiatric conditions or concomitant medications
4. Recent or unsubstantiated acute substance/alcohol abuse
5. Confounding or destabilizing factors that might obscure assessment
 - (a) Dominating extraneous physical or psychic symptoms
 - (b) Recent exposure to real or perceived traumatic events
 - (c) Unstable social, occupational, or living conditions
 - (d) Recent incarceration or hospitalization
6. Lack of clinical validity (relevance)
 - (a) Presenting symptoms lack sufficient impact on behavior and/or function (clinical relevance) to warrant treatment intervention
 - (b) Lack of a typical clinical presentation consistent with the known course of the disease (face validity)
7. Unreliable or unassessable subject (e.g., too disorganized, uncooperative, inconsistent) or unreliable or unavailable informant (caregiver)
8. Inadequate or insufficient rater interview (may lack documentation to warrant study eligibility)
9. A lack of a documented treatment history of response/non-response (in treatment resistant or partial treatment response studies)

A subject who is truly appropriate (valid) for enrollment in a clinical trial must have sufficient, measurable, and clinically relevant symptoms that go beyond the conventional checklists of the usual protocol inclusion and exclusion criteria [11,27,28]. The meaning of clinically relevant symptoms may vary somewhat depending on the nature of the study. In acute treatment studies, clinical relevance means that the identified acute symptoms are measurable *and* troublesome enough to warrant a new treatment intervention. Alternatively, in relapse prevention studies, patients are by definition relatively stable at the time of enrollment. In AD studies, the endpoint is generally focused on progressive deterioration in the absence of effective treatment. In studies that enroll

subjects who are relatively stable (relapse prevention) or likely to get worse without effective treatment (AD), it is important to document a *pre-treatment* history record that identifies relevant and measurable symptoms that *would* affect behavior or function *if* they recurred or got worse. In mild to moderate dementia studies, the presenting cognitive symptoms must have developed progressively over time, have some current impact on behavior or function, and be likely to progress further in the absence of an effective anti-dementia treatment. Recently, studies of prodromal AD have made subject selection by history less relevant because they seek patients *without* evidence of progressive decline or current functional impact and seek biomarkers of relevance instead.

The use of audio-digital recording of site-based interviews

We used an audio-digital pen recorder to record and review the site-based interviews. A computer tablet may also be used for this purpose although the pen is less obtrusive than a tablet and a more familiar object during an interview. The procedure is similar to the usual pen on paper interview and requires no additional props.

The site-based recording method does not require a separate, second interview because it simultaneously records the audio and written (digital) components of the actual site-based interview. The site-independent clinical rater is able to hear the interview and read the written notes to score symptoms by using *exactly the same data* obtained by the site-based rater. The dual assessments are based upon the *same* questions and the *same* subject responses given at the *same* time to the site-based rater. In addition, the process is less burdensome to the site than a second, remote interview and is not subject to the informational or temporal variance that would be generated by a separate interview.

The recording of site-based interviews is a form of quality assurance. The mere existence of an external surveillance system of this kind has a salutary effect on the quality of screening assessments. Independent review of the recordings can confirm that there was a complete and competent site-based interview done. Inadequate interviews lead to rater remediation and can improve performance and enhance the integrity of study data.

Beyond rating competency, the recordings permit entirely independent dual scoring of key rating instruments to assess scoring accuracy. The replication of site-based scores by a *blinded*, independent rater can not only affirm site-based ratings competency but also demonstrate rating precision [38].

The required recordings compel site-based interviewers to explore unanticipated validity issues that emerge during interviews that might have been skipped over without a

'recorded' surveillance process in place. These raters are fully aware that a potentially confounding issue that might obscure accurate assessment cannot be ignored because the dual reviewer may uncover the issue. For instance, the possibility of secondary gain as the motive to seek study entry (e.g., payments), or current substance abuse, if casually raised by the subject during the interview, must now be fully explored. Similarly, any recent traumatic experiences or acute stressors or abrupt life cycle changes that may have triggered the acute symptoms and might obscure accurate assessment during a trial must be clarified in order to justify subject eligibility.

The use of adjudication with the site investigator

Adjudication between the tier 2 reviewer and site-based investigator is required whenever an independent review of subject eligibility is challenged. Adjudication improves the precision of the subject selection process by inviting collaboration with the trial sites, providing remediation as needed, and reinforcing good clinical practice.

It is self-evident that an audio recording may not capture the full clinical picture or prevailing external circumstances for every potential subject. Therefore, telephonic adjudication with the site investigator is necessary and often useful for elaboration and clarification of historical and clinical information. The site-based investigator may be aware of additional information that warrants study eligibility. For instance, a subject may describe an *apparent* confounding factor (e.g., a recent move from one city to another or the illness of a parent) that may or may not be clinically relevant during the trial. Sometimes, the recorded site-based interview lacks sufficient information to verify the diagnosis, to confirm the presence or impact of important symptoms (e.g., delusions), or to evaluate the clinical relevance of recent changes in the subject's life. In the meta-analysis of the nine studies presented here, more than half of the tier 2 reviews were ultimately approved after telephone adjudication with the site investigator.

In our experience, telephone adjudication is a form of education and remediation. In the studies reviewed in this meta-analysis, the adjudication process had the interesting remedial effect of improving screen documentation for all submissions that followed adjudication discussions. As noted above, some investigators acknowledged that they increased their own level of eligibility diligence because of the rigor of the external review process.

Limitations of the C-VISA™ method

Two obvious limitations of the C-VISA™ method are (1) dependency on site-based rater competency and (2) reliance on audio-digital recordings as opposed to live interviews for the site-independent review.

Site-based ratings competency

It is obvious that the dual scoring method requires a competent site-based rater who is trained to conduct research interviews, administers complete interviews, and is willing to use an audio-digital recording device. The independent reviewer can query the site for more information but cannot interview the subject.

Rater training and certification conducted prior to a study does not necessarily equate with competent in-study interviews done with real study patients. The audio-recordings can identify incompetent or inadequate interviews and can lead to rater remediation. In most cases, site-based interviews provide sufficient data and yield high scoring correlations with the site-independent dual scores [39]. In some cases, the site-based rater cannot learn to conduct adequate interviews and must be replaced despite efforts at remediation.

Kobak and colleagues cautioned that unregulated site-based interviews might be inadequate and briefer than necessary to obtain sufficient clinical data [9]. In our experience, the requisite recording procedure and remediation efforts have actually *improved* the quality and integrity of the site-based interview and reinforced competent and full assessments.

C-VISA™ reliance on audio-digital recordings

In some studies, the audio-digital method of capturing historical and clinical information will not be sufficient to evaluate subject eligibility. For instance, an audio recording can ask questions but cannot capture the clinical observations made by a live interviewer about psychomotor retardation. A video recording may be needed in order to capture the visual appearance and behavioral responses that occur during the interview or to observe motor examinations (e.g., Parkinson's disease).

Neither audio- nor video-referenced assessments can fully capture the clinical nuance of a live face-to-face interview. For this reason, the adjudication discussion with the investigator is an essential part of the subject selection review process because it allows the independent reviewer to obtain more specific documentation and to clarify clinical information that may not be apparent on an audio or video recording (or during a telephone interview).

Non-specific factors affecting trial outcomes

Subject selection for CNS trials is a complex process that relies on the collection of accurate historical information and symptomatic data collected from an ill subject by a competent assessor. Clearly, there are limits to the precision of both diagnostic and symptom severity assessments for CNS diseases [11,28,40,41]. Non-specific, extraneous factors unrelated to the experimental drug treatment might affect subject behavior or performance and obscure accurate assessments at the screen visit and during the study as well.

It has been argued that non-specific factors unrelated to the candidate drug might account for more than 75% of the observed improvement in drug-treated groups in randomized clinical trials for MDD [42,43]. The C-VISA™ strategy is a systematic external review process that attempts to weed out some of these non-specific factors and reinforce data integrity. Although the C-VISA™ may reduce some of the manageable signal noise inherent in a double-blind clinical trial, it cannot eliminate all of the possible non-specific factors that are unrelated to drug treatment.

Summary and conclusions

The C-VISA™ process provides an unbiased, site-independent eligibility review strategy to confirm site-based assessments *prior* to randomization. As described in this meta-analysis, The C-VISA™ provided an independent clinical assessment of the recorded site-based screening interviews. The recording process eliminated the need for an entirely separate interview and was a cost-effective and time-efficient method for use in these clinical trials.

The C-VISA™ is essentially a quality assurance strategy that seeks to minimize the potential ‘noise’ that can adversely affect clinical trial outcomes. The recording method is scalable and can therefore be used in global, multi-national studies regardless of regional location or language.

The C-VISA™ external review method (1) identifies unreliable subjects and incompetent raters, (2) provides site-independent diagnostic verification, (3) provides symptom severity confirmation based upon dual scoring of recorded site-based interviews, (4) documents and affirms subject validity for the study, and (5) rules out the presence of comorbid conditions or confounding factors that might obscure the accurate assessment of symptoms during the study interval.

The C-VISA™ review process coupled with dual independent scoring of key rating instruments is a quality assurance strategy that provides a systematic site-independent eligibility filter to enhance the precision of subject selection and the integrity of study data. The process insists on appropriate documentation and a keener focus on accurate measurement of pre-randomization symptom severity. This customized audio-digital recording method is scalable and has been able to accommodate phase II and III global trials that use trial sites in Europe and Asia as well as the United States. The C-VISA™ strategy has broad applicability across the CNS spectrum because it achieves the objective of confirmatory site-independent review without producing excessive site or subject burden.

Abbreviations

AD: Alzheimer's disease; ADAS-cog: Alzheimer's disease assessment scale-cognitive component; BPRS: Brief Psychiatric Rating Scale; CDR: Clinical Dementia Rating Scale; CNS: central nervous system; C-VISA™: Clinical

Validation Inventory for Study Admission, the term C-VISA™ is protected by *Clintara LLC*, 2011; MDD: major depressive disorder; MINI: Mini International Psychiatric Inventory; PANSS: Positive and Negative Syndrome Scale; SAFER: validation criteria that examines state versus trait (S), assessability of symptoms (A), face validity (F), ecological validity (E), and the rule (R) of 3 Ps (persistence, pervasiveness, and pathology) related to current symptoms; SCID: Structured Clinical Interview for DSM diagnoses (versions IV or V); SD: standard deviation.

Competing interests

Steven D. Targum, M.D. has received consultation fees and/or vendor grants for rater training and surveillance services for investigational studies from Acadia Pharmaceuticals, Acumen, Alcobra, Alkermes Inc., AstraZeneca, BioMarin, BrainCells Inc., Civitas, Eli Lilly and Company, EnVivo (Forum) Pharmaceuticals, Euthymics, Forest Research, Functional Neuromodulation Inc., Johnson & Johnson PRD, Ironwood Pharmaceuticals, Methylation Sciences Inc., Mitsubishi Tanabe, NeoSync, Novartis Pharmaceuticals, Nupathe, Pfizer Inc., Prana Biotechnology Ltd., ReViva, Roche Labs, Saphiris, Sunovion, Takeda, Targacept, Theravance, and Transcept. He has equity interests in Clintara LLC, Methylation Sciences Inc., and Prana Biotechnology Ltd. J. Cara Pendergrass, Ph.D. is a full-time employee of Clintara LLC and received no fees or funding from any other institution of financial source.

Authors' contributions

Both SDT and JCP participated in the design of the customized C-VISA™ methodology, implementation of the program, analysis of results, and drafting of the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge Philip Rauh, Chelsea Toner, Alyssa Galley, Morgan Bailey, and Timothy Petersen for their participation and assistance in the implementation of the C-VISA™ program and analysis of the data. The term C-VISA™ is protected by *Clintara LLC*, 2011.

Received: 15 March 2014 Accepted: 16 July 2014

Published: 1 August 2014

References

1. Walsh BT, Seidman SN, Sysko R, Gould M: **Placebo response in studies of major depression: variable, substantial and growing.** *JAMA* 2002, **287**:1840–1847.
2. Khin NA, Chen YF, Yang Y, Yang P, Laughren TP: **Exploratory analyses of efficacy data from major depressive disorder trials submitted to the U.S. Food and Drug Administration in support of new drug applications.** *J Clin Psychiatry* 2011, **72**(4):464–472.
3. Khan A, Leventhal RM, Khan SR, Brown WA: **Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration Database.** *J Clin Psychopharmacology* 2002, **22**:40–45.
4. Khan A, Kolts RL, Rapaport MH, Krishnan KRR, Brodhead AE, Brown WA: **Magnitude of placebo response and drug-placebo differences across psychiatric disorders.** *Psychol Med* 2005, **35**:743–749.
5. Papakostas GI, Fava M: **Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD.** *Eur Neuropsychopharmacol* 2009, **19**:34–40.
6. Fava M, Evins A, Dorer D, Schoenfeld D: **The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach.** *Psychother Psychosom* 2003, **72**(3):115–127.
7. Kemp AS, Scholler NR, Kalali AH, Alphas L, Anand R, Awad G, Davidson M, Dubé S, Ereshefsky L, Gharabawi G, Leon AC, Lepine JP, Potkin SG, Vermeulen A: **What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it?** In *Proceedings of the First Collaborative Session between the ISCTM and the ISCD. Brussels, Belgium: Schizophrenia Bulletin*; 2008:2007.
8. Kobak KA, Feiger AD, Lipsitz JD: **Interview quality and signal detection in clinical trials.** *Am J Psychiatry* 2004, **162**:628.
9. Kobak KA, Kane JM, Thase ME, Nierenberg AA: **Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms?** *J Clin Psychopharmacol* 2007, **27**(1):1–5.
10. Targum SD: **Evaluating rater competency for CNS clinical trials.** *J Clin Psychopharm* 2006, **26**(3):308–310.

11. Targum SD, Pollack MH, Fava M: **Re-defining affective disorders: relevance for drug development.** *CNS Neurosci Ther* 2008, **14**:2–9.
12. Targum SD, Houser C, Northcutt J, Little LA, Cutler AJ, Walling DP: **A structured interview guide for global impressions: reliability and validity for CNS trials.** *Ann Gen Psychiatr* 2013, **12**:2–6.
13. Leon AC, Marzak PM: **More reliable outcome measures can reduce sample size requirements.** *Arch Gen Psychiatry* 1995, **52**:867–871.
14. Perkins DO, Wyatt RJ, Bartko JJ: **Penny wise and dollar foolish: the impact of measurement error on sample size requirements in clinical trials.** *Biol Psychiat* 2000, **47**:762–766.
15. Muller MJ, Szegedi A: **Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials.** *J Clin Psychopharmacol* 2002, **22**:318–325.
16. DeBrotta DJ, Demitrack MA, Landin R, Kobak KA, Greist JH, Potter WZ: **A comparison between interactive voice response system-administered HAM-D and clinician-administered HAM-D in patients with major depressive episode.** Boca Raton, FL: 39th Annual NCDEU Meeting; 1999.
17. Feltner DE, Kobak KA, Crockatt J, Haber H, Kavoussi R, Pande A, Greist JH: **Interactive voice response (IVR) for patient screening of anxiety in a clinical drug trial.** Phoenix, AZ: 41st Annual NCDEU Meeting; 2001.
18. Dunn J: **Comparison of site-based versus centralized ratings in a study of generalized anxiety disorder.** New Orleans, LA: 163rd Annual APA meeting; 2010.
19. Chandler G, Targum SD, Pollack M, Iosifescu D, Mischoulon D, Perlis R, Witte J, Fava M: **Validation of patients for a CNS trial of major depressive disorder.** Hollywood, FL: 49th Annual NCDEU meeting; 2009.
20. Schoemaker J, Gaur R, Chawla V, Jansen W, Szegedi A: **Expert rater assisted score evaluation (ERASE): a new method to enhance signal detection in randomized, placebo-controlled clinical trials.** Hollywood, FL: 49th Annual NCDEU; 2009.
21. Glaudin V, Smith WT, Ferguson JM, DuBoff EA, Rosenthal MH, Mee-Lee D: **Discriminating placebo and drug in generalized anxiety disorder (GAD) trials: single vs. multiple raters.** *Psychopharmacol Bull* 1994, **32**:175–178.
22. Simon GE, Revicki D, VonKorff M: **Telephone assessment of depression severity.** *J Psychiat Res* 1993, **27**(3):247–252.
23. Rohde P, Lewinsohn PM, Seeley JR: **Comparability of telephone and face-to-face interviews in assessing axis I and II disorders.** *Am J Psychiatr* 1997, **154**:1593–1598.
24. Kobak KA, Leuchter A, DeBrotta D, Engelhardt N, Williams JBW, Cook I, Leon A, Alpert J: **Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response.** *J Clin Psychopharmacol* 2010, **30**(2):193–197.
25. Moore HK, Wohlreich MM, Wilson MG, Mundt JC, Fava M, Mallinckrodt CH, Greist JH: **Using daily interactive voice response assessments to measure onset of symptom improvement with duloxetine.** *Psychiatry* 2007, **4**(3):30–38.
26. Shen J, Kobak KA, Zhao Y, Alexander M, Kane J: **Use of remote centralized raters via live 2-way video in a multicenter clinical trial for schizophrenia.** *J Clin Psychopharmacol* 2008, **28**(6):691–693.
27. Kendell R, Jablensky A: **Distinguishing between the validity and utility of psychiatric diagnoses.** *Am J Psychiatr* 2003, **160**(1):4–12. *Am J Psychiatr*, **126**: 983–987.
28. Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS: **Limitations of diagnostic and assessment instruments for mental disorders.** *Arch Gen Psychiatr* 1998, **55**:109–115.
29. First MB, Spitzer RL, Gibbon M, Williams JBW: **Structured Clinical Interview for DSM-IV Axis I Disorders (SCID).** New York Psychiatric Institute: Biometric Research Department; 1995.
30. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC: **The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10.** *J Clin Psychiatry* 1998, **59**(20):22–33.
31. Targum SD, Little JA, Lopez E, DeMartinis N, Rapaport M, Eshesky L: **Application of external review for subject selection in a schizophrenia trial.** *J Clin Psychopharmacol* 2012, **32**(2):825–826.
32. Kay SR, Fiszbein A, Opler LA: **The positive and negative syndrome scale (PANSS) for schizophrenia.** *Schizophrenia Bull* 1987, **13**:261–276.
33. Overall JE, Gorham DR: **The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling.** *Psychopharmacol Bull* 1988, **24**:97–99.
34. Folstein MF, Folstein SE, McHugh PR: **Mini-mental state. A practical method for grading the cognitive state of patients for the clinician** *J Psychiatric Res* 1975, **12**(3):189–198.
35. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL: **A new clinical scale for the staging of dementia.** *Brit J Psychiatr* 1982, **140**:566–572.
36. Rosen WG, Mohs RC, Davis KL: **A new rating scale for Alzheimer's disease.** *Am J Psychiatry* 1984, **141**(11):1356–1364.
37. Gaur R, Ramirez L, De Santi S, Schoemaker J: **Rater training on SANS and monitoring of rater performance during clinical trials.** In *ISCTM 5th Annual meeting*; 2009.
38. Leigh-Pemberton R, Memisglo A, Targum S, Pendergrass C, Rauh P, Marshall R, Silverman B, de Somer M, Ehrich E: **Blinded dual ratings confirm primary site-based ratings in an MDD trial.** Hollywood, FL: Annual ASCP Meeting; 2014.
39. Asgharnejad M, Targum S, Burch D, Gibertini M, Fava M: **Surveillance strategies to improve study outcomes in a depression study.** Boca Raton, FL: 52nd Annual NCDEU meeting; 2012.
40. Keller WR, Bernard A, Fischer BA, Carpenter WT Jr: **Revisiting the diagnosis of schizophrenia: where have we been and where are we going?** *CNS Neurosci Ther* 2011, **17**:83–88.
41. Kendler KS, Gardner CO Jr: **Boundaries of major depression: an evaluation of DSM-IV criteria.** *Am J Psychiatr* 1998, **155**:172–177.
42. Lambert MJ: *Handbook of psychotherapy integration.* New York: Basic Books; 1992:94–129.
43. Kirsch I, Sapirstein G: **Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication.** *Prevention and Treatment* 1998, **1**(2):Article 0002a. doi: 10.1037/1522-3736.1.1.12a.

doi:10.1186/s12991-014-0021-3

Cite this article as: Targum and Pendergrass: Site-independent confirmation of subject selection for CNS trials: 'dual' review using audio-digital recordings. *Annals of General Psychiatry* 2014 **13**:21.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

