Worldwide Clinical Trials

Practical Expertise in Neuroscience Clinical Research

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED • MEDICALLY DRIVEN

We are clinicians, scientists, and researchers who facilitate the development of life-changing medicines. Since 1986, our highly consultative approach has ensured that each drug development program receives personal consideration and commitment.

The medical and scientific heritage of our team is reflected in our staff and expertise in neuroscience disorders; in fact, over half of the studies we conduct are in neuroscience indications. Our methodological rigor and ability to link experimental design to operational imperatives is the force behind the world-class clinical trials we conduct globally.

Neuroscience Leadership at Worldwide Clinical Trials

Neal R. Cutler, M.D.

Chief Executive Officer

Dr. Cutler, CEO of Worldwide Clinical Trials, is a board-certified psychiatrist and is also board-qualified in both neurology and clinical pharmacology. Currently, he is president of the American Foundation for Clinical Pharmacology and serves on a special emphasis panel review committee for the National Institute on Aging for NIH. Dr. Cutler has also been instrumental in the design and clinical development of nearly 200 compounds in numerous therapeutic areas and has particular expertise in central nervous system disorders.

Michael Murphy, M.D., Ph.D.

Chief Medical and Scientific Officer

Dr. Murphy is board-certified in psychiatry and has a doctorate in pharmacology, with training at Tulane University, Stanford University and the Mt. Sinai School of Medicine. His professional career has spanned 25 years and his positions within the pharmaceutical industry emphasize the integration of medical and scientific acumen with operational excellence. His supervisory responsibilities as Chief Medical & Scientific Officer at WCT are international in scope and include the design and implementation of protocol feasibility assessments and protocol development for phases I–IV including non-interventional research.

As a faculty member within the Center for Experimental Pharmacology and Therapeutics at Harvard-MIT Division of Health Sciences and Technology, he has been a lecturer for 15 years within a competitive and credentialed clinical investigator training program.

Neuroscience Leadership

Henry J. Riordan, Ph.D.

Executive Vice President, Medical and Scientific Affairs

Dr. Riordan is currently responsible for the scientific conduct and service delivery of all neuroscience clinical research initiatives undertaken by WCT. Dr. Riordan specializes in CNS clinical trials methodology and is a licensed psychologist who has published over 100 peer-reviewed abstracts, articles, books and book chapters focusing predominantly on innovative neuroscience trial methods. Dr. Riordan has been involved in the assessment, treatment and investigation of various CNS disorders in both industry and academia for the past 20 years. He has been the primary author of numerous protocols and several clinical development plans across a variety of neurologic, psychiatric and analgesic indications. Dr Riordan spent several years in the departments of psychiatry and neurology at Thomas Jefferson, Dartmouth, University of Pennsylvania and Stony Brook Medical Schools where he acquired advanced training in biostatistics, experimental design, neurophysiology, neuroimaging and clinical neuropsychology.

Tomislav Babić, MD, PhD

Vice President, Neuroscience

Dr. Babic is a board certified neurologist and Affiliate Professor of Clinical Neurology. At Worldwide Clinical Trials, he is responsible for the scientific and medical leadership of global neurology clinical research initiatives. This includes aspects of hypothesis generation and testing, protocol/strategic program design and development, as well as assistance in the analysis and clinical interpretation or results for all phases of clinical drug development.

Dr. Babic, a therapeutic leader in neurology medical and scientific affairs, has designed protocols and programs for randomized controlled clinical trials in populations with early and advanced Parkinson's disease, Alzheimer's disease, multiple sclerosis, epilepsy, stroke, migraine, and neurodegenerative disorders, implementing the up-to-date evidence-based science in clinical drug research and development.

Neuroscience Leadership

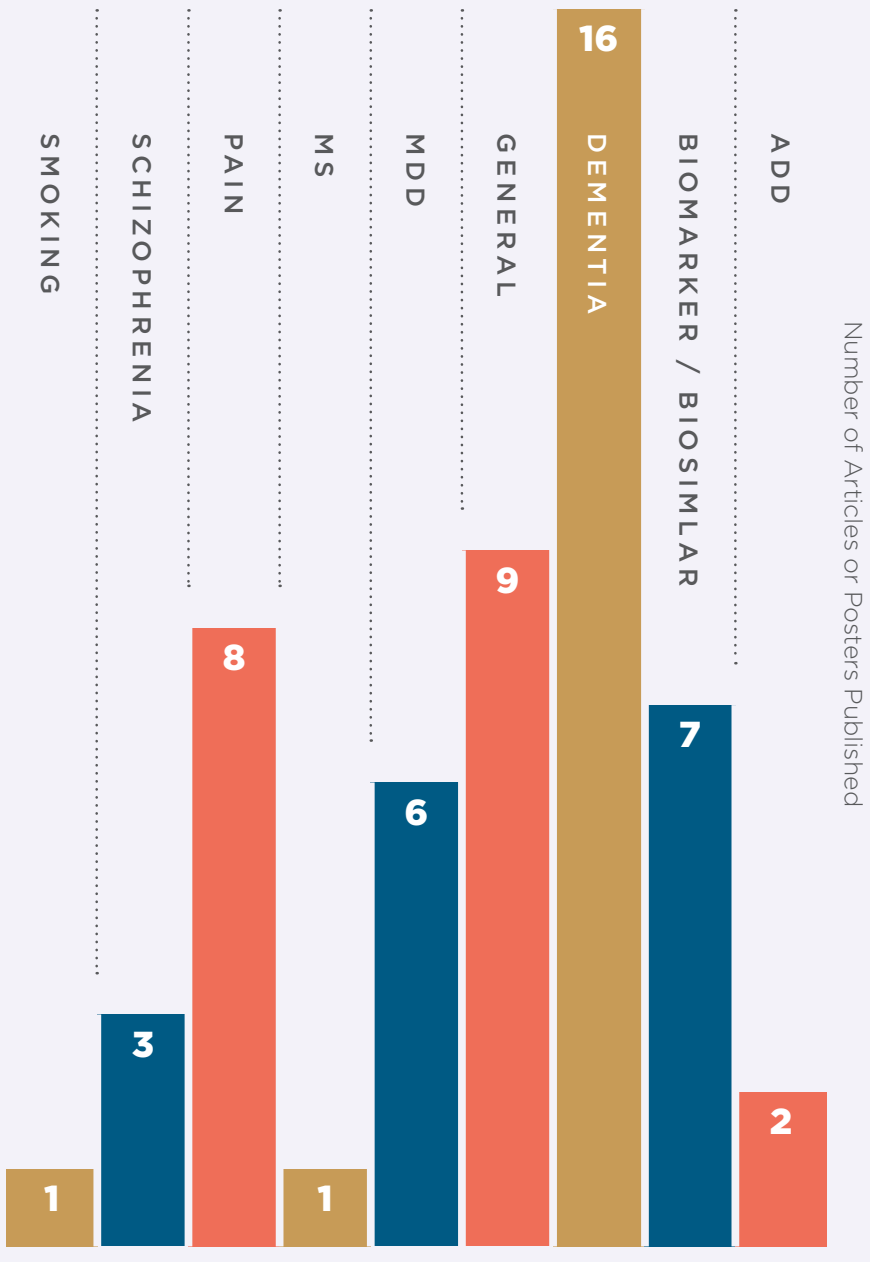
Douglas Lytle, PhD, MBA

Executive Director, Clinical Analytics, Training and Surveillance (CATS)

Dr Douglas Lytle, has a PhD in neuro/psychopharmacology from University of North Texas Health Science Center (1995) and an MBA from Cornell University (1998). He has over 20 years' of CRO experience in project management, business development, clinical operations, finance, and rater reliability services. He has served as an Adjunct Instructor for medical and graduate programs at University of North Texas Health Science Center, Villanova University, and Widener University.

Dr Lytle heads up the Clinical Analytics, Training and Surveillance (CATS) group within WCT. This independent department consists of 17+project management and clinical staff members who are dedicated to specialized rater training and surveillance of clinical trial sites - focusing on decreasing the variability of subjective assessments associated with CNS indications.

WCT CNS Expertise Presented & Published



Did you know?

Our staff has contributed to the development of every drug class applicable to Alzheimer's disease in the past 20 years.

CATS

Worldwide Clinical Trials' Clinical Analytics, Training and Surveillance (CATS) group is fully integrated with our study teams — helping to enhance signal detection by improving the quality of sites, of patients, and of the outcomes data.

A COMPARISON OF RATING QUALITY BETWEEN SUBTITLED AND SPOKEN HAMILTON DEPRESSION RATING SCALE INTERVIEWS

Bethanne Friedmann, PsyD[1], Henry Riordan, PhD[2], Evan Braxton[1], Christopher J. Weber, PhD[1], Michael F. Murphy, MD, PhD[1], Neal R. Cutler, MD[3]

[1]Worldwide Clinical Trials, King of Prussia, PA, [2]CEDRA Clinical Research, San Antonio, TX, [3]Worldwide Clinical Trials, Beverly Hills, CA

ABSTRACT

Expert raters help ensure accurate signal detection by minimizing inter-rater variability and reducing rater bias. This study compares the benefits of utilizing recorded interviews conducted in Russian to train primary Russian speakers versus English subtitles of these interviews to train English speaking raters. This method permits examination of subtitles in isolation, and compares rating concordance between subtitled versus spoken Hamilton Depression Rating Scale (HAMD-17)[1] interviews in the two language rating groups.

BACKGROUND

Global expansion of clinical trials has created the demand for high quality expert raters across a number of different countries. Many rater training programs utilize interviews in spoken English with subtitling provided in the primary/native language to train raters.

Despite the shift toward multinational clinical trials, most studies to date have utilized interviews in spoken English with subtitling provided in primary (local) language to train raters, without knowing the effectiveness of subtitles versus the spoken, native language.

METHODS

Two HAMD-17 interviews were conducted in Russian, and given English subtitles, using the Structured Interview Guide for the HAMD-17 (SIGH-D)2, to test 2 groups of raters. Group 1 (n=32) were Russian (RN) raters whose primary language was Russian (and did not speak English); Group 2 (n=13) were United States (US) raters whose primary language was English (and did not speak Russian). Didactic training was given before interviews in each group's native language, and interviews were shown in the same order for both groups. US raters rated interviews based on English subtitles. Demographic data was collected from all raters including education, clinical, research experience, and HAMD experience.

Figure 1.

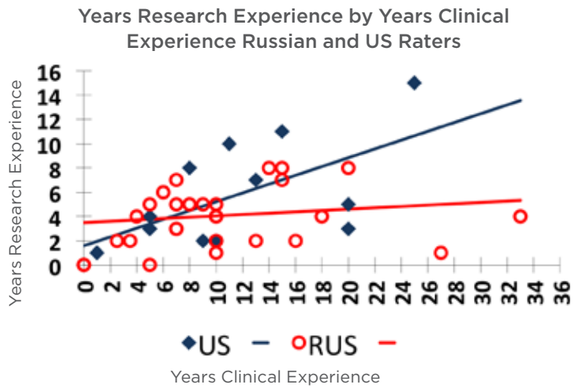


Table 1.

	HAMD-17 Scores			
	Interview 1		Interview 2	
	Average	STDEV	Average	STDEV
US raters (n = 13)	29.0	4.14	35.0	1.29
Russian raters (n = 32)	28.5	2.57	34.1	2.41
	p=0.625		p= 0.224	

RESULTS

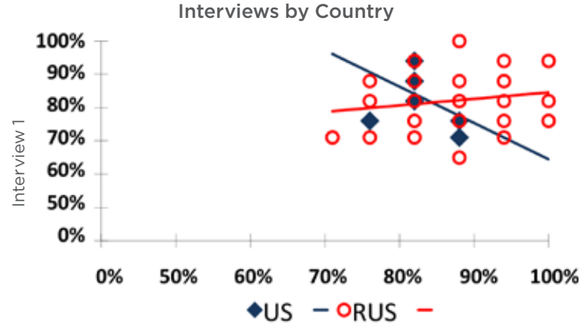
Gold standards were established through item-by-item review of each recorded interview to determine rating quality. In addition, rater concordance levels were used to determine homogeneity in perception of observed symptomatology (higher concordance = lower Standard Deviations).

- There was no significant difference in average HAMD-17 score between the RN and US raters on either interview.
- RN and US raters both showed high overall concordance rates of 88% and 83%, respectively, with the gold standard ratings (p=0.796).
- Demographic variables did not appear to affect overall concordance.
- Individual item analysis yielded statistically significant differences between rater groups in SD's.
- Standard Deviations (SDs) of the US raters was significantly smaller on the second interview than the first (p=0.224), suggesting practice effects when utilizing subtitles (both groups improved with practice).
- There was very little dispersion (characterized by low SDs) whether interviews are in native language or subtitled.
- Cultural differences between rater groups were reflected by individual HAMD-17 item differences but this did not affect overall score concordance.

CONCLUSIONS

- This study confirms that adequate training methodology using appropriate language and subtitling can overcome any a priori differences in ratings due to regional differences, cultural biases and local medical practice.
- There were no significant differences between Russian and US raters across multiple dependent variables with both groups benefiting from practice and didactic feedback.
- Variables that affect rating accuracy such as number of years experience, type of degree, and various demographic variables should be evaluated carefully when choosing raters for future depression studies.

Figure 2.



ACKNOWLEDGEMENTS

1 Hamilton, M., 1960 A rating scale for depression. Journal of Neurology, Neurosurgery, and Psychiatry 23, 56-62.

2 Williams, J.B., 1988 A structured interview guide for the Hamilton Depression Rating Scale. Archives of General Psychiatry 45, 742–747.

No conflicts of interest exist in the research and development of this poster

THE EFFECT OF SIMULTANEOUS TRANSLATIONS AND EXPERIENCE ON POSITIVE AND NEGATIVE SYNDROME SCALE (PANSS) SCORING PERFORMANCE FOR EASTERN EUROPEAN RATERS

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Bethanne Friedmann, PsyD[1], Henry Riordan, PhD[2], Evan Braxton[1], Christopher J. Weber, PhD[1], Michael F. Murphy, MD, PhD[1], Neal R. Cutler, MD[3]

[1]Worldwide Clinical Trials, King of Prussia, PA, [2]CEDRA Clinical Research, San Antonio, TX, [3]Worldwide Clinical Trials, Beverly Hills, CA

ABSTRACT

Twenty-five raters from 19 sites in Russia, Ukraine, Croatia and Serbia participated in a didactic PANSS1 training session utilizing an audience response system. Rater performance was correlated with level of prior experience. Raters completed a brief review before scoring an initial patient video (pre-test condition), then received a thorough item by item review of the PANSS scoring by an expert presenter. Raters were shown a second patient video (post-test condition). Both video-recorded patient interviews were subtitled, and raters received transcripts and translations of the SCI-PANSS2. Raters were allowed to discuss scores and ask questions in their local language after the first video. Changes in average item deviations from the gold standards were calculated for each rater for each video and analyzed according to experience levels. Following training all raters were well within apriori benchmarks developed by the independent expert consensus panel, and showed an non-significant decrease in deviations of PANSS positive items from pre-test to post-test conditions (t=1.15, p = .26). When raters were divided into two groups according to number of years of experience rating the PANSS (1-5 years, n = 12; 6-plus years, n = 13), average item deviations of the 6-plus years group dropped significantly to zero. Demographic variables including number of prior trials, scale frequency, research and clinical experience were also assessed. Number of prior trials and clinical experience had no significant impact on training benefits. A thorough didactic training session including simultaneous translation of the PANSS with local language discussion can produce rater performance exceeding predefined benchmarks. Benefits of training were especially apparent for raters with higher levels of PANSS and research experience (>5 years), suggesting the usefulness of experienced Eastern European sites in conducting global schizophrenia trials.

BACKGROUND

It is becoming standard for clinical studies evaluating efficacy of new anti-psychotic agents to be conducted in multiple countries prior to receiving FDA and/EMEA approval. According to ClinicalTrials.gov3, there are currently 449 active clinical studies worldwide in schizophrenia. Over half of these studies are conducted in countries where English is not the primary language. Nearly one- third of all studies do not include English-speaking countries and yet the majority of investigators' meetings and training materials are delivered in English. The goal of the investigators' meeting and rater training is to ensure that communications regarding assessments and study-specific conventions are standardized across sites worldwide. Most rater training programs utilize didactic training on scoring conventions and scoring of patient video interviews to align raters.

The PANSS is a 30-item assessment utilized to measure symptoms of schizophrenia in clinical studies. It is divided into 7 positive, 7 negative and 16 general items associated with the symptoms of schizophrenia. For this study the 7 positive items of the PANSS were chosen for analysis since these are the most sensitive to change. PANSS items analyzed were:

- **P1.** Delusions
- **P2.** Conceptual disorganization
- **P3.** Hallucinatory behavior
- **P4.** Excitement
- **P5.** Grandiosity
- **P6.** Suspiciousness/persecution
- **P7.** Hostility.

Table 1. Average Experience

AVERAGE NUMBER OF TRIALS	AVERAGE RESEARCH EXPERIENCE (YEARS)	AVERAGE CLINICAL EXPERIENCE (YEARS)
3	7	13

Table 2. PANSS Experience

EXPERIENCE WITH SCALE (YEARS)	PANSS	CGI	NUMBER OF ADMINISTRATIONS (WITHIN THE PAST 2 YEARS)	PANSS	CGI
1-2	1	1	11-15	3	3
2-3	2	2	16-25	3	
3-5	9	10	26-50	8	6
6-10	8	9	50+	11	16
10+	5	3			

METHODS

Twenty-five raters from 19 sites in 4 countries, Russia, Ukraine, Croatia and Serbia, participated in a PANSS1 training and certification program during an investigators' meeting in Europe. At the investigators' meeting, the PANSS didactic training was simultaneously translated into the 4 languages. Two complete PANSS patient video interviews were subtitled and the raters were provided with translated transcripts and the translations of the SCI-PANSS2.

A consensus panel was conducted prior to the investigator meeting to set a gold standard for the scores of both patient video interviews. The consensus panel consisted of 10 raters independent of this study; 5 US experts; 2 Russian experts; 2 Serbia Experts and 1 Ukrainian expert. All had at least 5 years of experience scoring the PANSS in clinical studies and advanced degrees. Nine members of the consensus panel were doctoral level physicians or psychologists. The 10th member was a doctoral level student with over 5 years of experience and who had conducted the PANSS patient video interviews.

Prior to scoring the first PANSS patient video interview, there was a brief overview of the PANSS scoring conventions presented in English with simultaneous translation. After scoring the first patient video interview, raters were given an item by item review of all 30-items including anchor level scoring justifications as determined by the expert consensus panel. The scoring discussion was facilitated through the use of an audience response system. Raters were able to discuss scores and ask questions in local language via simultaneous translation. After a thorough review of the first patient video interview, raters were shown a second patient video interview to score. The scores of the second patient video interview were collected without group discussion to be used for rater certification.

The demographic information of all 25 raters was assessed to evaluate potential impact on scoring of the positive items. Experience was evaluated for number of trials, PANSS usage, research, clinical and PANSS experience. (see Table 1 and 2) The 25 raters were divided by experience into groups, low (1-5 years) and hi (6 plus years). Lastly, each of the 7 positive items were evaluated for significant deviations on the pre and post patient video interviews as well as change between pre and post on individual items (see Figure 1).

Figure 1. P-Item Differences

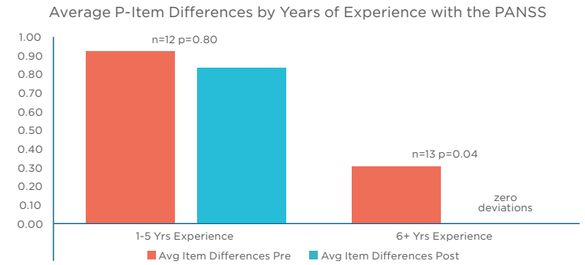


Table 3. Experience Analysis

MEDIAN SPLIT TWO TAILED T-TEST P VALUES					
	NUMBER OF TRIALS	RESEARCH EXPERIENCE (YEARS)	CLINICAL EXPERIENCE (YEARS)	PANSS FREQUENCY (2 YRS)	PANSS EXPERIENCE (YEARS)
LOW GROUP	P=.72	P=.81	P=.81	P=.43	P=.80
HIGH GROUP	P=.26	P=.04	P=.04	P=.50	P=.04

Table 4. Pre Test Score Deviations

PRE TEST SCORE DEVIATIONS RELATIVE TO GOLD STANDARD (HI GROUP, LOW GROUP) N = 25							
Points From Range	P1	P2	P3	P4	P5	P6	P7
-1	0,2					2,2	
0	23	20	25	23	24	20	25
1		1,4		0,2	0,1	1,0	

Table 5. Post Test Score Deviations

POST TEST SCORE DEVIATIONS RELATIVE TO GOLD STANDARD (HI GROUP, LOW GROUP) N = 25							
Points From Range	P1	P2	P3	P4	P5	P6	P7
-1	0,3			0,2			
0	21	24	24	21	25	25	25
1	0,1	0,1	0,1	0,2			

RESULTS

25 raters were divided into two groups by years of experience rating the PANSS (1-5 years, n = 12 and 6-plus years n=13)(see Table 3). Both groups showed improvement concordance on the combined positive items for the post patient video interview scoring, however a two tailed t-tests showed that there was significant improvement in concordance for the higher PANSS experience group in the combined and individual positive item scoring. The individual item analysis on positive items showed the most variability on P2, Conceptual disorganization and P6, Suspiciousness/persecution. On the pre patient video interview item P2, 5 out of 25 (20%) raters scored the item higher than the consensus panel. On the pre patient video interview item P6, 4 out of 25 (16%) raters scored the item lower and 1 out of 25 (4%) scored the item higher than the consensus panel. (See Table 4) On the post patient video interview there was less variation amongst raters. On P2, 1 rater (4%) scored the item lower and on P6, all raters (100%) scored in concordance with the consensus panel. (See Table 5 and Table 6).

CONCLUSIONS

A thorough didactic training session including simultaneous translation of the PANSS with discussion in local languages may have contributed to the improved performance of all raters. This suggests that simultaneous translation should be considered for multinational clinical trial to increase understanding and alignment of raters. This may result in greater efficacy of multinational clinical studies.

- Simultaneous translation of the didactic review along with practice scoring and a thorough review of all 30 PANSS items prior to scoring the certification patient video interview may have improved the concordance of raters with the expert consensus panel
- The expert consensus panel should be representational of the regions covered to take into consideration cultural differences and biases
- It was observed that the more experienced raters showed more significant improvement on the positive items from pre to post patient video interview which may indicate that they incorporated the lessons of the training more readily than the less experienced raters

Table 6. Scoring Analysis

	PRE			POST			OVERALL PRE TEST VS POST TEST	OVERALL HI PRE VS HI PST	OVERALL LO PRE VS LO PST
	HI	LO	P-Value	HI	LO	P-Value	P-Value	P-Value	P-Value
P1	0.00	0.15	0.1356	0.00	0.31	0.5596	0.0000	0.000	0.000
P2	0.08	0.31	0.2220	0.00	0.08	0.5900	0.0253	0.007	0.557
P3	0.00	0.00	0.9150	0.00	0.08	0.3079	0.0424	0.070	0.294
P4	0.00	0.15	0.9749	0.00	0.31	1.0000	0.0000	0.000	0.000
P5	0.00	0.08	0.4475	0.00	0.00	N/A	N/A	0.000	0.000
P6	0.25	0.15	0.6206	0.00	0.00	0.8767	0.0000	0.000	0.000
P7	0.00	0.00	0.5619	0.00	0.00	0.8767	0.0000	0.000	0.000

REFERENCES

1 – Kay S, Opler LA, Fiszbein A: Positive and negative syndrome scale: Technical manual. MHS. 2006.

2 – Opler LA, Kay SR, Lindenmayer JP & Fiszbein A: Strutured clinical interview – Positive and negative syndrome scale. MHS. 1992.

3 – http://clinicaltrials.gov/ct2/results/map?term=schizophrenia&recr=Open&map=NS retrieved June 2009.

INTERACTIVE VOICE RESPONSE SYSTEM FOR PATIENT REPORTED CHANGES IN MOOD STATE USING OUTBOUND CALL PROMPTING

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Christopher J Weber, PhD[1], Henry Riordan, PhD[2], Evan Braxton[1], Bethanne Friedmann, PsyD[1], Mark Leibowitz, MD[2], Neal R Cutler, MD[3]

[1]Worldwide Clinical Trials, King of Prussia, PA, [2]CEDRA Clinical Research, San Antonio, TX, [3]Worldwide Clinical Trials, Beverly Hills, CA

ABSTRACT

Background: The Profile of Mood States (POMS) 1 has been widely used in a variety of clinical trials to identify and quantify affective states2. Despite its use, little research has examined the POMS as administered via Interactive Voice Response (IVR). We examined the psychometric properties of the IVR POMS and compliance associated with an outbound calling procedure.

Methods: Thirty-one healthy normal subjects from a phase II trial completed the IVR POMS daily. Outpatient POMS ratings were completed daily following a random call to the subject between 9am to 7pm. Inpatient POMS ratings were completed at three equally spaced time intervals throughout the day.

Results: Raw scores were analyzed from the two weeks prior to inpatient (pre-inpatient), the two weeks inpatient, and the two weeks immediately following inpatient (post-inpatient). Total Mood Disturbance (TMD) score, Tension-Anxiety (T-A), Depression-Dejection (D-D), and Anger-Hostility (A-H) subscale scores were averaged for each subject as well. Although there was some suggestion for more symptom endorsement during the inpatient period, one-way ANOVA's showed no significant differences among the POMS scores when compared across the three time periods, suggesting relatively stable mood across in/outpatient status, as well as the relative utility of once daily versus multiple administrations. Although completion rates from pre-inpatient (80%) to post-inpatient (65%) showed a significant decrease (t= -3.23 p<.003), this decrease did not appear to be related to POMS total or subscale scores as a median split resulting in "high" vs. "low" compliance groups showed no significant differences in TMD or subscale scores for either pre or post-inpatient conditions.

Conclusions: IVR administration of the POMS is an effective data collection method for inpatient and outpatient trials. Daily administration is as efficient as three times, suggesting that subjects accurately summate their mood state over 24 hours. Compliance with the IVR POMS appears to wane over time for reasons unrelated to mood states suggesting the need for additional prompting as the trial progresses.

INTRODUCTION

- The Profile of Mood States (POMS) consists of a list of 65 adjectives rated individually on a 5-point scale: not at all, a little, moderately, quite a bit, and extremely.
- The validity of the POMS has been well-established through previous publications, particularly in studying mood variations in normal adult populations and non-psychotic subjects.
- Interactive Voice Response (IVR) technology, accessed via telephone networks and using touch-tone keypad selection, enhances assessment methodology by increasing efficiency of data collection and minimizing patient completion errors.
- An IVR version of the POMS has not been previous published.

OBJECTIVE

- The objective of this study was to examine the psychometric properties of the IVR POMS and compliance associated with an outbound calling procedure.

SUBJECTS

- 87 subjects (32 females and 55 males) were selected for this analysis.
- Subjects were between 18 and 75 years of age. Females =34.9 years (range= 20-72), Males =34.5 years (range=19-60).
- Subjects were current cigarette smokers scoring >5 on the Fagerström Test for Nicotine Dependence.
- Subjects had no current or past history of major Axis I or II disorder as determined by the Structured Clinical Interview for DSM-IV (SCID).
- Subjects had no current or past history of suicidal ideation or suicidal behavior as determined by the Columbia Suicide Severity Rating Scale (C-SSRS).

METHODS AND DESIGN

- Study enrollment was part of a single-site, 1:1 randomized, double-blind, placebo-controlled study with a 12-week treatment period for smoking cessation. The design included a two-week outpatient baseline period, two weeks of inpatient treatment, and outpatient treatment for the remaining weeks.
- Subjects were enrolled in sequential cohorts of 15-20 every four weeks.
- Subjects completed the IVR POMS in both outpatient and inpatient phases of the study.
- As outpatients, subjects completed the IVR POMS once daily. Each subject provided a personal contact telephone number for the IVR system to initiate outbound calls. These automated calls served as a prompt for subjects to complete their daily IVR POMS, and were randomly generated between 9am and 9pm to collect diverse sampling of their mood data. Subjects could not complete their IVR POMS prior to this outbound call. Furthermore, to increase compliance, automated "reminder" calls were made to subjects who did not complete their IVR POMS within one hour of receiving their initial automated telephone call.
- As inpatients, IVR POMS were completed three times daily, at equally spaced, fixed intervals.
- As illustrated in the figure below, data analysis for this research for restricted to two weeks before inpatient, two weeks inpatient, and two weeks immediately following inpatient.

- The 65-item POMS was adapted to IVR, as each item was read to the patient, prompting their immediate numeric response corresponding to the anchors descriptors. Subjects were allowed to correct any incorrect responses.
- Subjects were provided detailed written instructions for IVR completion of the POMS.
- POMS total score and subscale scores examined for this analysis were:
  - Total Mood Disturbance (TMD; all 65 items)
  - Three selected subscales:

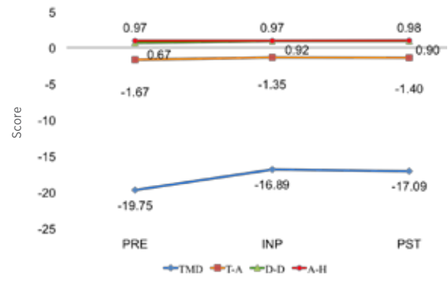
Anger/Hostility (A-H)		Tension/Anxiety (T-A)		Depression/Dejecjeon (D-D)	
Angrey	Ready to Fight	Tense	Relaxed	Unhappy	Lonely
Peeved	Rebellious	Shaky	Uneasy	Sorry	Miserable
Grouchy	Deceived	On edge	Restless	Sad	Gloomy
Spiteful	Furious	Panicky	Nervous	Blue	Desparate
Annoyed	Bad-tempered		Anxious	Hopeless	Helpless
Resentful				Unworthy	Worthless
Bitter				Discouraged	Terrified
					Guilty

- IVR POMS total and subscale scores were automatically computed in real time.
- A unique feature of the WCT IVR system was the ability to generate alerts via email directly to site staff when:
  a) any subjects' individual total or subscale score exceeded a pre-determined threshold and/or
  b) the subject was not compliant in completing their POMS
- All incoming email alerts were reviewed within 48 hours (most were reviewed the next day), and action taken by site staff ranged from ongoing patient monitoring or brief contact with the subject, to psychiatric review by the PI or designee.

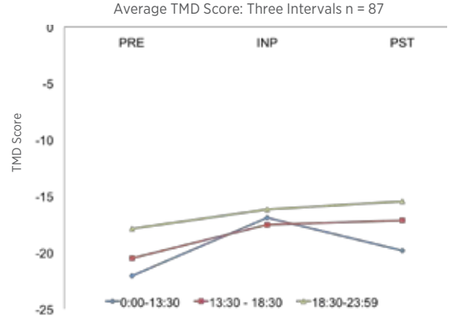
RESULTS

- Total Mood Disturbance (TMD) score, Tension-Anxiety (T-A), Depression-Dejection (D-D), and Anger-Hostility (A-H) subscale scores were averaged for each subject. One-way ANOVA's showed no significant differences among the POMS scores when compared across the three time periods.

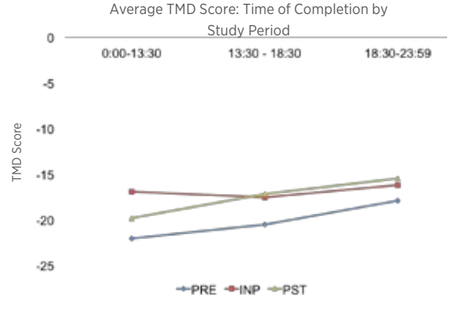
TMD and Subscales, Averages Across Three Stages



Average TMD Score: Three Intervals n = 87



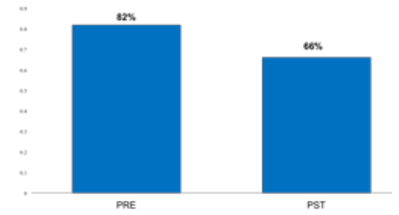
Average TMD Score: Time of Completion by Study Period



- A median split of TMD scores (n=851) in study period by time of completion and vice versa.
- No statistical differences in either analysis across all groups. However, most groups hinted towards more symptomology later in the day.

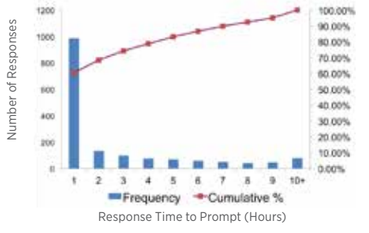
RESULTS (Continued)

Overall Compliance Rates Pre and Post inpatient Period

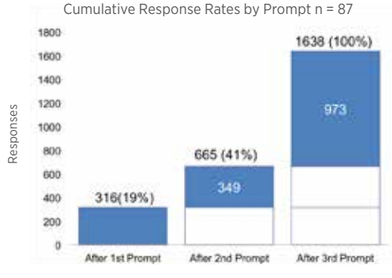


- Overall compliance with IVR POMS was 82% and 66% in Pre-Inpatient and Post-Inpatient, respectively.
- Furthermore, of all POMS completions, figure (left below)shows that 60% of subjects completed their IVR POMS within one hour of receiving the first outbound reminder call. 80% of IVR POMS were completed within four hours of the first prompt.
- Figure (below right) illustrates the effectiveness of the repeated outbound reminder calls.

Distribution of Response times with Cumulative %



Cumulative Response Rates by Prompt n = 87



CONCLUSIONS

- IVR administration of the POMS is an effective data collection method for inpatient and outpatient clinical trials.
- IVR POMS shows relatively stable mood across in/outpatient status, as well as the relative utility of once daily versus multiple administrations. In other words, daily administration is as efficient as three times daily, suggesting that subjects accurately summate their mood state over 24 hours.
- Compliance with IVR POMS appears to decrease with time, but the cause of this decrease is unrelated to POMS scores. This suggests the need for additional prompting as the trial progresses.

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

UNDERSTANDING SITE PREFERENCES WITH eCOA TECHNOLOGIES USED TO INCREASE THE RELIABILITY OF CLINICAL ASSESSMENTS: A MULTI-NATIONAL SURVEY

Douglas Lytle, Ph.D., MBA[1], Elan A. Cohen, Ph.D.[1], Colleen R. Rock, BA[1], Andrew E. Komorowsky, MS[1], Bethanne Friedmann, Psy.D.[1], Michael Murphy, M.D., PhD[1]

[1]Worldwide Clinical Trials, King of Prussia, PA

ABSTRACT

Introduction: As technology continues to grow, so does its use within the clinical trial industry. In particular, tablet and computer-based eSource/e-Clinical Outcome Assessments (eCOA) solutions are being used more and more as tools to increase the reliability of clinical assessments. While research indicate these technologies improve the efficiency and quality of rater administered and patient reported outcome assessments compared to paper instruments (Tiplady, 2014; Williams et al., 2015), little empirical exploration has been conducted examining the experiences of those who use these technologies within trials on a daily basis. The current study investigates study coordinator and raters' perceptions on their use of a variety of technologies being applied within the industry. Method: Site coordinators and raters were anonymously surveyed from US and ROW sites. The sites designated to receive the survey had previously participated in numerous psychiatric and neurocognitive studies. The site staff were queried about their experiences using these technologies as well as various demographic information. Conclusion: Obtaining the experiential realities of site staff who utilize assessment technologies is critical to increasing the use and acceptance of the technologies. The goal of this investigation is to better understand site perceptions regarding these technologies in an effort to address potential shortcomings, leading to their greater usage that will enhance the overall quality of rater assessments

BACKGROUND

eSource/e-Clinical Outcome Assessments (eCOA) are increasingly utilized in the collection of patient data as part of clinical trials. Laptop/tablet programs, smartphone/web-based applications and interactive voice systems (IVRS) are among the most commonly used today. Tiplady et al., (2014) demonstrated that developed technologies have shown to be effective in reduction of study costs and increase the reliability of outcome data and inclusion of appropriate subjects. However, there has been limited focus on site perceptions regarding these technologies. The authors of the current study sought to investigate site preferences in using the newer technology of eSource/eCOA, as there has been an increase in its usage for conducting clinical assessments.

DESIGN

Approximately 1500 web-based surveys were designed and distributed by Worldwide Clinical Trials via email in order to inform the utilization of paper- based versus electronic-based data acquisition platforms. All data are proprietary to WCT. Responses were limited to one per computer.

The focus of the questionnaire was to compare differences between eCOA technologies and traditional paper assessments through five questions that were presented in English via SurveyMonkey. The first four questions focused on background information: study role, location, years of clinical trial experience, and number of previous studies using eCOA solutions. Respondents were allowed to choose "Rater" and/or "Coordinator" to define their role. Respondents who indicated both roles were excluded to more clearly compare perceptions based on different roles.

The fifth question focused on preference of various eCOA technologies based on a 5-point Likert scale (Strongly Dislike, Dislike, Neutral, Prefer and Strongly Prefer). For analysis, "Strongly Dislike" and "Dislike" responses were combined into an overall "Dislike" category; "Strongly Prefer" and "Prefer" responses were also combined into one category. The overall percent of respondents that preferred a technology (Strongly Preferred and Preferred) was subtracted from the overall percent of respondents that Disliked (Strongly Disliked and Disliked) the technology. The difference was defined as a Preference Ratio and is represented in the difference of the percentages in overall basis points; positive numbers indicate stronger preferences whereas negative numbers indicate stronger "dislike" partialities.

A final question was open for respondents to provide free text comments on their preferred method for conducting assessments.

On some analyses as defined below, we combined the testing methods of tablet with stylus, tablet with keyboard, and laptop as single eCOA category; the digital pen was kept as a separate eCOA category.

For analysis purposes, responses were transformed from categorical ranges (years of experience, number of studies using eCOA, etc.) to the average of the available range. For example, if the range for years of experience was 1-5, 6- 10, etc. these answers were converted to the average of the range to be 3 or 8 years of experience, respectively.

Chi square distribution and ANOVA testing of the results were included as described below, which occasionally included multiple comparisons of a single dataset. Because there were no prior assumptions about the expected pattern of outcomes, a decision was made to control for multiple comparisons using the Bonferonni correction (dividing p value by the number of multiple comparisons).

RESULTS

A total of 319 respondents completed the survey with over half of the respondents from North America (51%), followed by Western Europe, Asia-Pac, Eastern EU, and Russia/Ukraine as depicted in the table 1 below.

Table 1: Regional Demographics of Responders

Answer Options	Response Percent	Response Count
Asian-PAC	15.7%	50
Eastern EU	7.5%	24
North America	50.8%	162
Russia / Ukraine	1.6%	5
Western EU	21.9%	70
	Total Responses	319

Years of clinical trial experience was skewed, with 60% of respondents indicating 10 or less. Nearly 25% of respondents had over 16 years of clinical trial experience as shown in Table 2. There was a good representation and range of experience with eCOA solutions from the responders as depicted in Table 3

Table 2: Number of Years of Experience of Responders

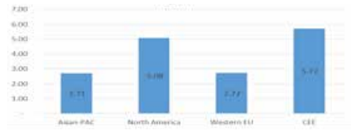
Answer Options	Response Percent	Response Count
1-5 Years of Experience	32.6%	104
6-10 Years of Experience	27.9%	89
11-15 Years of Experience	15.7%	50
16-20 Years of Experience	8.5%	27
> 20 Years of Experience	15.4%	49
	Total Responses	319

Table 3: Number of studies you have used a tablet / electronic device to conduct assessments?

Answer Options	Response Percent	Response Count
0	14.7%	47
1-3	33.5%	107
4-6	22.9%	73
7-9	7.2%	23
> 9	20.7%	66
N/A	0.9%	3
	Total Responses	319

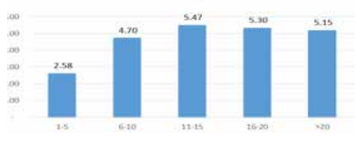
Surprisingly, sites based in CEE averaged the most number of studies using eCOA, followed by sites based in North America as depicted in the Figure 1 below. An ANOVA across all data indicated a significant difference among regions regarding the rater experience with eCOA solutions ($p<0.00000$). No difference between North America and Eastern EU was indicated ($p=0.29$) and significant differences between North American and Western EU ($p<0.0000$) and Asia Pac ($p<0.0000$) were identified.

Figure 1: Average Number of eCOA Studies by Region



Not surprising, a rater or coordinator with overall less experience was shown to also have overall less experience with eCOA, but surprisingly there was not a stronger trend (Figure 2). There was a slight trend for more eCOA experience with more years of experience with an R value of 0.765, but the trend was not significant.

Figure 2: Average Number of eCOA Studies by Years of Experience



RESULTS (Continued)

The key question in the survey asked "Please rate each of the following methods for conducting assessments based on your experience using each of the below technologies." Respondents to this question indicated that paper based assessments were highly favored, with a 62 basis point Favorability Ratio (67% favored paper, 6% did not favor paper) as shown in Table 4. The most disliked method was Digital pen recorder with a -31 point favorability ratio (12% preferred, 43% did not favor). The tablet with stylus was more preferred than tablet with keyboard whereas a laptop was favored over either tablet options.

Table 4: Overall Preference of eCOA Options Compared to Paper Assessments

Answer Options	Strongly Dislike	Dislike	Neutral	Prefer	Strongly Prefer	% Prefer	% Dislike	Preference Ratio
Tablet with Stylus	33	40	91	68	34	32%	23%	9%
Tablet with Keyboard	28	53	93	58	27	27%	26%	1%
Laptop	21	44	102	69	41	35%	21%	14%
Audio/Digital Pen Recorder	65	70	82	29	8	12%	43%	-31%
Paper Assessment	4	14	80	108	105	67%	6%	62%

When this question was analyzed based on the role of the respondent, Coordinators consistently liked eCOA more than the raters as shown in Tables 5 and 6 below. Since the data from "Table with Stylus", "Tablet with Keyboard", and "Laptop" tended to be similar across Tables 4, 5 and 6, we combined this data into one "eCOA" preference ratio; this ratio did not include the responses from "Digital Pen Recorder" as the responses were markedly different from the other eCOA options. The Coordinators had a modest 17 point preference ratio for the combined eCOA solutions, whereas Raters had a 0 basis point preference for eCOA solutions (data not shown in table). Nevertheless, paper assessments were still highly favored in both groups, with Coordinators preferring paper assessments (52 point preference ratio) a little less than Raters (69 point preference ratio).

Table 5: Coordinator Preference of eCOA Options Compared to Paper Assessments

Answer Options	Strongly Dislike	Dislike	Neutral	Prefer	Strongly Prefer	% Prefer	% Dislike	Preference Ratio
Tablet with Stylus	5	12	25	17	13	42%	24%	18%
Tablet with Keyboard	5	10	54	12	11	25%	16%	9%
Laptop	3	11	30	19	14	43%	18%	25%
Audio/Digital Pen Recorder	15	18	23	8	3	16%	49%	-33%
Paper Assessment	2	4	27	29	19	59%	7%	52%

Table 6: Rater Preference of eCOA Options Compared to Paper Assessments

Answer Options	Strongly Dislike	Dislike	Neutral	Prefer	Strongly Prefer	% Prefer	% Dislike	Preference Ratio
Tablet with Stylus	18	19	40	33	14	38%	30%	8%
Tablet with Keyboard	18	32	85	25	9	20%	30%	-9%
Laptop	11	25	50	28	15	33%	28%	5%
Audio/Digital Pen Recorder	32	41	41	14	0	11%	57%	-46%
Paper Assessment	0	5	38	52	59	72%	3%	69%

When the preference data was analyzed based on region, CEE had the strongest preference for paper assessments (82 point preference ratio), followed by North America (70), Asia-PAC (51) and Western EU (50) (Table 7). When tablet and laptop preferences were combined, North America and CEE had the least favorable ratings (-3 preference points), with North America (1), Western EU (13) and Asia-PAC (20) having higher preference ratios for overall eCOA solutions.

RESULTS (Continued)

Table 7: Overall eCOA Preference vs Paper Preference Ratio by Region

Region	Overall eCOA Preference Ratio	Paper Preference Ratio
Asian-PAC	20%	51%
CEE	-3%	82%
North America	1%	70%
Western EU	13%	50%

Those with zero experience with tablet/eCOA solutions had the highest preference for tablet/eCOA based solutions (31 preference points). Those with any experience with eCOA solutions consistently had higher preference for paper as compared to those respondents with no experience with eCOA. (Table 8).

Table 8: Overall eCOA Preference vs Paper Preference Ratio by Number of eCOA Studies

Number of eCOA Studies	Overall eCOA Preference Ratio	Paper Preference Ratio
0	31%	46%
1-3	6%	70%
4-6	11%	57%
>7	-4%	73%

DISCUSSION

It is understood that the industry is clearly moving from paper based solutions to electronic solutions. The transition to eCOA solutions may be slow just as it was for EDC where it took 10+ years for the vast majority of new studies to adopt EDC over double-data entry of paper based CRF forms. Over these years, many of the issues and complications of EDC have been resolved - or accepted as a necessary challenge that is outweighed by the benefits of EDC. It is clear that eCOA solutions will continue to grow in use and acceptability over the coming years as we address some of the challenges and issues with eCOA solutions.

As we analyzed the data, some interesting trends appeared. For example, Study coordinators liked the eCOA technologies more than raters; this may reflect how technology makes the lives of the coordinator easier, whereas raters may be experiencing more difficulty with implementing the technology directly with the patient. Indeed, many of the comments from respondents were concerned that the physical barrier and challenges with the eCOA solution when interviewing the patient may affect the reliability of data derived from eCOA.

Not surprisingly, a rater with overall less experience was also shown to have overall less experience with eCOA, but there was not a stronger trend with more years of experience. The combination of lack of overall experience and lack of eCOA experience within this less experienced group may be reason to question the validity of these perceptions as they may not have a strong experiential history to develop well-formed opinions re paper or eCOA based assessments. For example those with zero experience with tablet/eCOA solutions actually had the highest preference for tablet/eCOA based solutions (31 preference points); a clear demonstration that "grass is greener on the other side of the fence."

One of the primary goals of this survey was to understand sites perceptions of eCOA solutions in order to: 1) potentially address problems, challenges and issues associated with eCOA options, 2) increase the use of tablet-based eCOA solutions, and 3) ultimately increase reliability of outcomes data through greater site acceptance of eCOA technologies. Based on our experience - and free-text entries from the respondents - some of the eCOA issues that should be considered when implementing eCOA solutions include: easier interface that is less distracting for the clinician to utilize while interviewing a subject, ensure that the eCOA solution is not developing a barrier between the clinician and the subject, provide easier set up for each subject, ensure font is large enough on eCOA, ensure strong tech support, ensure backup/ contingency plans regarding the device are appropriately addressed, ensure training and orientation to the site staff is comprehensive. Throughout the industry, we should continue to ask sites opinions regarding their perceptions and how we can best implement eCOA and other technology solutions. By removing issues and obstacles, we will ultimately increase the reliability of the assessments as well as make the day-to- day life/study activities/work easier at the site. Of course, site, sponsors, vendors and patients all win when the reliability of outcomes data increases in CNS indications, ultimately leading to more drugs being approved to treat the manifold unmet needs of these patient populations.

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

VALIDITY OF SEMANTIC MEMORY TEST AS PREDICTOR OF CHANGE IN ADAS-COG SCORES IN MCI

Erin B. Kornsey, MS[1], Bethanne Friedmann, PsyD[1], Michael F. Murphy, MD, PhD[1], Neal R. Cutler, MD[2]
[1]Worldwide Clinical Trials, King of Prussia, PA, [2]Worldwide Clinical Trials, Beverly Hills, CA

ABSTRACT

There are many clinical trials evaluating disease modifying agents in subjects with mild cognitive impairment (MCI). This poster examines the predictive value of a neuropsychiatric test used to characterize the cognitive deficit in MCI in comparison to a well validated assessment commonly used in Alzheimer's disease trials. ROC analysis was used to compare the neuropsychiatric tests. Preliminary statistical analyses from a multinational trial suggest the Free and Cued Selective Recall Reminding test does not reliably predict short-term longitudinal change in the Alzheimer's disease Assessment Scale-cognitive subscale in the MCI population.

ABSTRACT

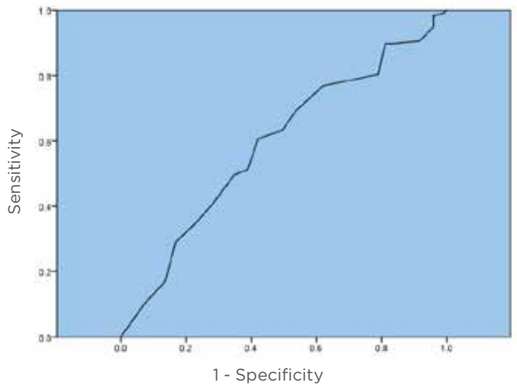
Neuropsychiatric test batteries are utilized to measure cognitive performance in many indications. In a multinational study for MCI a neuropsychiatric test battery included a novel assessment for some sites, the Free and Cued Selective Recall Reminding test (FCSRT)[1]. This reminding test for verbal episodic memory is an adaptation of the Grober & Buschke paradigm to diagnose MCI phenotype[2]. The FCSRT test is considered a very sensitive and specific assessment for the diagnosis of prodromal Alzheimer's Dementia (AD). This test may distinguish early AD from subtypes of MCI were conversion to AD is less common. For the purposes of this clinical trial, the FCSRT was administered for diagnostic purposes and was not utilized as an outcome measure. The inclusion criteria were free recall score < 20 and a total recall score of < 40. The focus of this review was the ability of the FCSRT to track short-term changes in the Alzheimer's disease Assessment Scale-cognitive subscale (ADAS-cog).

METHODS

The analyses were conducted on subjects who completed a screening FCSRT, qualified for randomization and subsequently provided an ADAS-cog at baseline and at week 24. Descriptive statistics were generated on the FCSRT. Change scores were computed for baseline ADAS-cog and week 24 resulting in a variable of increase or decrease in ADAS-cog. The change score for each patient was compared to the baseline FCSRT's free recall score and total recall score. Receiver operating characteristics (ROC curves) were generated on the change score variable of the ADAS-cog and free recall score and total score of the FCSRT using all patients randomized, regardless of treatment group assignment. ROC analysis tests for sensitivity and specificity of models based on the theory of signal detection. An excellent model would have an area under the curve of .9 – 1 while a test that was a poor model would result in an area under the curve of .5-.6.

Figure 1.

ROC Curve of Free Recall Scores and ADAS-cog change



RESULTS

The ADAS-cog scores for 93 subjects increased at week 24, while the scores for 117 subjects decreased. (See Table 1 for a summary of scores on free recall and total recall by change score of ADAS-cog). In this study, the FCSRT for either free recall and total recall appeared to be poor predictors of subsequent ADAS-cog performance. The area under the curve for the free recall scores was .588 (see Figure 1 for ROC), and the area under the curve for the total recall score of FCSRT was .606 (Table 2)

Table 1. Means (Standard Deviations) of neuropsychiatric assessments.

	Free Recall Score	Total Recall Score
Increase in ADAS-cog in 24 weeks (n=93)	11.60 (5.26)	28.52 (8.65)
Decrease in ADAS-cog in 24 weeks (n=117)	13.24 (5.24)	31.23 (8.80)

Table 2. Comparing ROC curves.

	Area Under the Curve	Standard Error
Free Recall	.588	.039
Total Recall	.606	.039

CONCLUSIONS

A cross-sectional, screening FCSRT failed to mirror subsequent change in ADAS-cog scores over 24 weeks for this multinational MCI trial. A limitation of the above study is that the FCSRT was only used as a screening measure and was not administered at subsequent weeks so that so that correlations in change could be directly linked. Also, the time reference was limited to 24 weeks, which may be inadequate to demonstrate changes in either measure. Please note these preliminary analsyes were generated using blinded data. Further analysis should include a logistic regression to see if each instrument contributes to a subject's decreased cognitive abilities.

REFERENCES

1 Sarazin, M. et al. (2007) Amnestic syndrome of medical temporal types indentifies prodromal AD: A longitudinal study. Neurology, 69, 1859-67.

2 Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. Developmental Neuropsychology, 3, 13-36

No conflicts of interest exist in the research and development of this poster

INTERNATIONAL: +44 (0) 20 7121 6161

AMERICAS: +1 610 964 2000

worldwide.com

WORLDWIDE CLINICAL TRIALS

SCIENTIFICALLY MINDED • MEDICALLY DRIVEN