Worldwide Clinical Trials
Practical Expertise in Neuroscience Clinical Research
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.
Dr 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.
Our staff has contributed to the development of every drug class applicable to Alzheimer’s disease in the past 20 years.

Did you know?

**WCT CNS Expertise Presented & Published**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Articles/Posters Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>3</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>8</td>
</tr>
<tr>
<td>Pain</td>
<td>6</td>
</tr>
<tr>
<td>MS</td>
<td>9</td>
</tr>
<tr>
<td>MDD</td>
<td>16</td>
</tr>
<tr>
<td>General</td>
<td>7</td>
</tr>
<tr>
<td>Dementia</td>
<td>16</td>
</tr>
<tr>
<td>Biomarker / Biosimilar</td>
<td>2</td>
</tr>
</tbody>
</table>

**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\textsuperscript{1}, Henry Riordan, PhD\textsuperscript{2}, Evan Braxton\textsuperscript{1}, Christopher J. Weber, PhD\textsuperscript{1}, Michael F. Murphy, MD, PhD\textsuperscript{1}, Neal R. Cutler, MD\textsuperscript{3}

\textsuperscript{1}Worldwide Clinical Trials, King of Prussia, PA, \textsuperscript{2}CEDRA Clinical Research, San Antonio, TX, \textsuperscript{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 (HAM-D-17) 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), 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.

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.

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

<table>
<thead>
<tr>
<th>Interview</th>
<th>RN (n = 32)</th>
<th>Average</th>
<th>STDEV</th>
<th>RN (n = 13)</th>
<th>Average</th>
<th>STDEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>US raters</td>
<td>28.5</td>
<td>2.57</td>
<td>34.1</td>
<td>2.41</td>
<td>0.629</td>
<td></td>
</tr>
<tr>
<td>Russian rates</td>
<td>28.5</td>
<td>2.57</td>
<td>34.1</td>
<td>2.41</td>
<td>0.629</td>
<td>0.224</td>
</tr>
</tbody>
</table>

Table 1.
BACKGROUND

It is becoming standard for clinical studies evaluating efficacy of new antipsychotic agents to be conducted in multiple countries prior to receiving FDA and EMEA approval. According to ClinicalTrials.gov, 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 are conducted in countries where English is not the primary language. Nearly 1,000 active clinical studies worldwide in schizophrenia. Over half of these studies and research experience (>5 years), suggesting the usefulness of a thorough didactic training session including 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.

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 translations should be considered for multinational clinical trials to improve understanding and alignment of raters. This may result in greater efficacy of multinational clinical studies.

RESULTS

Twenty-five raters were divided into two groups by years of experience rating the PANSS; n = 12 (3–5 plus years) and n = 13 (6 years plus). When raters were divided into two groups according to number of years of experience rating the PANSS (n = 1, 6–plus years) 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 prior to rating in the PANSS with local language discussion may result in greater efficacy of multinational clinical studies.

SYNDROME SCALE (PANSS) SCORING PERFORMANCE FOR EASTERN EUROPEAN RATERS

METHODS

Twenty-five raters from 19 sites in Russia, Ukraine, Croatia and Serbia participated in a didactic PANSS5 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 performance for Eastern European raters. The expert consensus panel should be representational of the regions covered to take into consideration cultural differences and biases may result in greater efficacy of multinational clinical studies.

It was observed that the more experienced raters showed significantly improved concordance of raters, and the less experienced raters showed improvement concordance on the combined positive items for the post-test patient video interview which may indicate that they incorporated the lessons of the training more readily than the less experienced raters.

REFERENCES

BACKGROUND: The Profile of Mood States (POMS) has been widely used in various clinical trials to identify and quantify affective states. Despite its use, little research has examined the POMS as administered via an Interactive Voice Response (IVR) system. We aimed to determine whether the psychometric properties of the IVR POMS are comparable to those of the paper-and-pencil version.

METHODS: Thirty-two healthy adults were selected for this analysis. Outpatient POMS ratings were completed daily for 14 days, followed by an inpatient period of 2 weeks. IVR POMS ratings were completed daily, at equally spaced, fixed intervals. 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. The reliability of the POMS has been assessed through various methods, including the Intraclass Correlation Coefficient (ICC). In this study, the ICC for interrater reliability was calculated for both the IVR and paper-and-pencil versions of the POMS.

RESULTS: The intraclass correlation coefficient for interrater reliability for the IVR and paper-and-pencil versions was 0.92 for the Total Mood Disturbance (TMD) score, 0.90 for the Tension-Anxiety (T-A) subscale, 0.89 for the Depression-Dejection (D-D) subscale, and 0.91 for the Anger-Hostility (A-H) subscale. These results suggest that the IVR POMS is a reliable tool for assessing mood states.

CONCLUSIONS: The IVR POMS shows relatively stable mood across inpatient and outpatient status, as well as the relative utility of daily versus multiple administrations. Although completion rates from pre-inpatient to post-inpatient were high, there were some indications of decreased compliance in the post-inpatient period. This suggests the need for additional prompting as the trial progresses.

ABSTRACT

Background: The Profile of Mood States (POMS) has been widely used in a variety of clinical trials to identify and quantify affective states. Despite its use, little research has examined the POMS as administered via an Interactive Voice Response (IVR) system. We aimed to determine whether the psychometric properties of the IVR POMS are comparable to those of the paper-and-pencil version.

Methods: Thirty-two healthy adults were selected for this analysis. Outpatient POMS ratings were completed daily for 14 days, followed by an inpatient period of 2 weeks. IVR POMS ratings were completed daily, at equally spaced, fixed intervals. 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. The reliability of the POMS has been assessed through various methods, including the Intraclass Correlation Coefficient (ICC). In this study, the ICC for interrater reliability was calculated for both the IVR and paper-and-pencil versions of the POMS.

RESULTS: The intraclass correlation coefficient for interrater reliability for the IVR and paper-and-pencil versions was 0.92 for the Total Mood Disturbance (TMD) score, 0.90 for the Tension-Anxiety (T-A) subscale, 0.89 for the Depression-Dejection (D-D) subscale, and 0.91 for the Anger-Hostility (A-H) subscale. These results suggest that the IVR POMS is a reliable tool for assessing mood states.

CONCLUSIONS: The IVR POMS shows relatively stable mood across inpatient and outpatient status, as well as the relative utility of daily versus multiple administrations. Although completion rates from pre-inpatient to post-inpatient were high, there were some indications of decreased compliance in the post-inpatient period. This suggests the need for additional prompting as the trial progresses.
抽象

介绍了技术发展如何带来，以及我们因缺少临床试验数据而带来的，可能对临床决策有影响的风险。eCOA系统可以实时收集数据，并且在试验过程中进行更多质量控制，从而提高临床试验的效果。然而，eCOA的使用也存在一些挑战，如数据安全和隐私保护问题，以及如何在多中心临床试验中整合这些系统。本文旨在了解研究者对eCOA技术的接受程度和偏好。

结果

通过网络问卷调查收集了来自北美的170名研究者的数据。调查内容包括研究者的偏好、对eCOA技术的接受程度以及对这些技术的使用经验。结果表明，大多数研究者对eCOA技术持积极态度，且在使用经验方面表现出了明显的地区差异。

设计

本次调查采用网络问卷方式，设计了多中心临床试验，通过收集研究者的背景信息、对eCOA技术的偏好以及对eCOA技术使用经验的评估来了解研究者的偏好和接受程度。

讨论

虽然eCOA系统在临床试验中的使用越来越普遍，但仍有挑战需要解决。研究者对eCOA技术的偏好和接受程度不同，可能影响了eCOA技术在临床试验中的使用。未来的研究应进一步探索这些差异的原因，并提出相应的解决方案，以促进eCOA技术在临床试验中的更广泛使用。
VALIDITY OF SEMANTIC MEMORY TEST AS PREDICTOR OF CHANGE IN ADAS-COG SCORES IN MCI
Erin B. Kornsey, MS1, Bethanne Friedmann, PsyD1, Michael F. Murphy, MD, PhD1, Neal R. Cutler, MD2
1Worldwide Clinical Trials, King of Prussia, PA, 2Worldwide 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). This reminding test for verbal episodic memory is an adaptation of the Grober & Buschke paradigm to diagnose MCI phenotype. 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.

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.

<table>
<thead>
<tr>
<th></th>
<th>Free Recall Score</th>
<th>Total Recall Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=93)</td>
<td>11.60 (5.26)</td>
<td>28.52 (8.65)</td>
</tr>
<tr>
<td>Decrease in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=117)</td>
<td>13.24 (5.24)</td>
<td>31.23 (8.80)</td>
</tr>
</tbody>
</table>

Table 2. Comparing ROC curves.

<table>
<thead>
<tr>
<th></th>
<th>Area Under the Curve</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall</td>
<td>.588</td>
<td>.039</td>
</tr>
<tr>
<td>Total Recall</td>
<td>.606</td>
<td>.039</td>
</tr>
</tbody>
</table>

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

Figure 1. ROC Curve of Free Recall Scores and ADAS-cog change

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

<table>
<thead>
<tr>
<th></th>
<th>Free Recall Score</th>
<th>Total Recall Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=93)</td>
<td>11.60 (5.26)</td>
<td>28.52 (8.65)</td>
</tr>
<tr>
<td>Decrease in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=117)</td>
<td>13.24 (5.24)</td>
<td>31.23 (8.80)</td>
</tr>
</tbody>
</table>

Table 2. Comparing ROC curves.

<table>
<thead>
<tr>
<th></th>
<th>Area Under the Curve</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall</td>
<td>.588</td>
<td>.039</td>
</tr>
<tr>
<td>Total Recall</td>
<td>.606</td>
<td>.039</td>
</tr>
</tbody>
</table>

Figure 1. ROC Curve of Free Recall Scores and ADAS-cog change

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

<table>
<thead>
<tr>
<th></th>
<th>Free Recall Score</th>
<th>Total Recall Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=93)</td>
<td>11.60 (5.26)</td>
<td>28.52 (8.65)</td>
</tr>
<tr>
<td>Decrease in ADAS-cog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in 24 weeks (n=117)</td>
<td>13.24 (5.24)</td>
<td>31.23 (8.80)</td>
</tr>
</tbody>
</table>

Table 2. Comparing ROC curves.

<table>
<thead>
<tr>
<th></th>
<th>Area Under the Curve</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall</td>
<td>.588</td>
<td>.039</td>
</tr>
<tr>
<td>Total Recall</td>
<td>.606</td>
<td>.039</td>
</tr>
</tbody>
</table>